PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 16 April 15, 1930 Number 4

THE VELOCITIES OF B-TYPE STARS

By Jan Schilt

Yale University Observatory

Communicated March 10, 1930

In his discussion of the radial velocities of faint B stars observed with the Victoria telescope, J. S. Plaskett[1] has represented the radial velocities by the formula

$$v = V \cos \lambda + K + rA \sin 2(l - l_0) \cos^2 b. \qquad (1)$$

The largest term in this expression is the first one, which represents the component of the sun's motion with respect to the stars. λ is the angular distance of a star from the anti-apex of the solar motion.

Since the B stars are distributed in a narrow zone along the galactic equator, the radial velocities give very little information about the component of the sun's motion in a direction at right angles to the galactic plane.

We must therefore restrict ourselves to the component in this plane and equation (1) may be written:

$$v = V_1 \cos (l - l_1) + K + rA \sin 2(l - l_0) \cos^2 b. \qquad (2)$$

In this equation we have

v = observed radial velocity,
V_1 = component of the solar motion in the galactic plane,
l = galactic longitude,
b = galactic latitude,
l_1 = longitude of the anti-apex, and
K = Campbell's constant.

The last term of the equation is the so-called rotational term which has been introduced by Oort in an attempt to explain certain irregularities on the basis of a dynamical theory put forward by Lindblad and further developed by Oort:

r = average distance of the group of stars,
A = constant of the rotational effect, and
l_0 = direction of the adopted center of rotation.

The values for V and the position of the apex are subject to considerable variation depending on the stars included in the solution. Since the individual motions of the B stars are small (say of the order of 14 km. per second), it is evident that a sample of 100 stars picked out at random should give a fairly representative determination of the solar motion in the galactic plane. In fact, the uncertainty of such a determination would be $50^{-1/2}$ or $1/7$ of the individual motions, or 2 km. per second corresponding to 6° in the position of the apex.

In various determinations of the solar apex for all different kinds of stars, based on radial velocities or proper motions, the declinations of the apices found differ more than the right ascensions. As a consequence of the orientation of the equator with respect to the galaxy the declination of the apex depends chiefly on the galactic longitude and this is just the quantity that can be determined for the B stars with high weight. Strömberg[2] has found that the declinations of the apex for different subdivisions of stars vary from 0° to 60° and for the stars only between spectral divisions A9 to B0 from +13° to +41°. Similar discrepancies have been found by many authors for stars of different spectral types and for subclasses of B stars and for stars of different magnitudes. As to the non-B stars a part of the anomalies is probably due to the presence of stars of high velocities. The individual velocities for these stars show a larger dispersion, which may account for part of the irregularities. For the B stars, however, according to the well-known small dispersion in the velocities, a very high stability in different solutions might be expected. The only explanation that this expectation is not fulfilled is that the mean error of the solar motion as computed in the illustration above, based on a sample of 100 stars, is not right and this is only possible if we sacrifice the tacit assumption that the accuracy of a determination increases with the square-root of the number of stars. This in turn means that the velocities of the B stars should not be considered as independent quantities. As soon as one realizes that the B stars according to their motions form a limited number of families, the members of each family sharing in one common motion, the theory of probability is no longer contradicted by the observed facts.

Even in the mean of a relatively small number of stars belonging to one family the small individual motions cancel out, and this mean value represents the group motion. Additional observations of stars belonging to the same family will increase the accuracy of the group motion but will not materially increase the weight of a determination of the sun's velocity

with respect to the B stars at large. Instead of the individual radial velocities the group motions enter as independent quantities in our determination.

It is well known that B stars show a strong tendency to clustering into groups which are sometimes referred to as "B star clouds." The parallelism in the proper motions of the B stars in one cloud was first shown by Kapteyn. It is almost impossible to escape from the conclusion that the clustering tendency must be linked with group motions. The problem of finding the laws which exist between the velocities of stars and their distribution is now in its infancy and will doubtless afford one of the most important fields of research for the years to come.

The number of bright clusters of B stars is small; it depends a little on whether one regards coherent smaller clusters as belonging together or not. Three to five star clouds contain the bulk of the bright B stars. The average group velocity is not accurately known but is probably as large as 10 km. per second. If so, various determinations based on bright B stars may differ up to 10 km. per second. For the same reason it is evident that the term K in (1) is different from zero.

If our sun occupied a place beyond the Orion group we might expect a negative K term unless it were cancelled by other families still further away.

Plaskett in the paper quoted has realized the importance of the motions for the significance of the K term. I agree fully that the day in which this term is to be looked upon as something physical should be regarded as past. In accordance with the recognition of the families it is to be expected that the K term will be different for different regions of the galaxy. In fact, this has been found by Freundlich and v. d. Pahlen.[3]

In order to obtain a more stable solution for the solar motion improvement can only be expected by considering more families, that is, by including fainter stars. It would be in agreement with the principle here developed if we find that the K term for the fainter B stars tends toward zero.

Plaskett arrives at the conclusion that the faint B stars in fact do not show a K term. I have, however, a serious objection to the method by which this result is obtained. He has taken the sun's velocity for granted. After the elimination of the component of an adopted solar velocity he enters the residual velocities in the equation (2). A solution is then made for the constant K and the second harmonic term in l. Plaskett contends that the circumstance that his stars cover only about 180° of the galactic circle is unfavorable for a solution of the solar motion, that is, the first harmonic term in l. If this is true it is still truer for the K term. The values of K and A, in fact, depend largely on the assumed solar velocity for the given distribution of the stars. In view of what is known about the

determinations of solar velocity there is not much reason to take this quantity for granted. It seems more logical to represent the observations by an ordinary solution and to make sure whether the residuals then found justify the introduction of a second harmonic. I have first made such a solution for the B3 to B5 stars tabulated in Plaskett's table II.

The solutions for the parallactic drift and K term according to the equation:

$$\rho = S \sin l + C \cos l + K,$$

where S and C represent the components of the correction to the adopted solar motion, gives

$$\begin{aligned} S &= -5.6 \text{ km./sec.} \pm 1.1 \text{ (mean error)} \\ C &= +0.7 \qquad\quad \pm 0.8 \text{ (mean error)} \\ K &= +3.6 \qquad\quad \pm 0.7 \text{ (mean error).} \end{aligned}$$

In view of the fact that velocities for neighboring stars should not be regarded as independent, unit weight has been given to Plaskett's mean values, as tabulated in his table II, irrespective of the number of stars except for three groups where the number of stars is less than 10, and these have been given weight 0.5.

The stars of magnitudes 5.5 to 6.5 and 6.5 to 7.5 do not seem to justify separate solutions; they have been entered all in one solution, the number of constants thus being three whereas Plaskett has four.

The differences $O-C'$ from the present solution are compared with Plaskett's residuals $O-C$ in the table.

1	$O-C'$ KM./SEC.	$O-C$ KM./SEC.	1	$O-C'$ KM./SEC.	$O-C$ KM./SEC.
22	+1.8	+0.7	148	−0.9	−0.6
29	−0.6	−2.7	164	+4.8	+3.0
37	+1.2	+0.8	167	−3.6	−4.2
56	−1.4	−3.2	177	+0.5	−0.8
56	+2.6	+2.5	182	0.0	−0.2
76	−2.4	−3.1	208	−1.5	+0.1
76	−2.4	−1.0	245	+1.0	+7.9
93	−4.0	−1.6	307	−3.0	+3.7
120	+2.7	+1.7	355	−0.5	+2.0
126	+2.4	+3.4	359	+0.6	+1.4

I find that the sum of the squares is 105 in my solution and 163 in Plaskett's. The number of residuals as large as 3.0 km./sec. is four against seven, and the distribution of signs shows not the slightest indication of a second-order term.

If the solar velocity is assumed to be 20 km./sec. toward the apex 271°, +28°, which value Plaskett has adopted in M. N. R. A. S., **88,** 399, the galactic components are:

$$X = -17.4 \text{ km. per second}$$
$$Y = -\ 6.9 \text{ km. per second}$$
$$Z = -\ 7.0 \text{ km. per second.}$$

The corrections found for the faint B stars change these values to

$$X = -16.7 \text{ km. per second}$$
$$E = -12.5 \text{ km. per second}$$
$$Z = -\ 7.0 \text{ km. per second,}$$

corresponding to a solar velocity of 22 km./sec. toward the apex 280°, +40°.

It is interesting to note that the change is exactly in the direction found by Strömberg and others for stars of nearly the same spectral types.

It then appears that a slight modification in the adopted value of the solar motion enables us to represent the observations more satisfactorily than can be done by the introduction of a second harmonic term.

The K term, +3.6, is of the usual order of magnitude.

The stars of spectral types B0 to B2 in Plaskett's table II show a very different behavior. The residuals appear to fit a second harmonic term very well, and their motions must be related to those of the c stars and the O stars and the brighter B0–B2 stars, which show the same effect. This, in my opinion, is due to the fact that the stars of these classes exhibit group motions of certain galactic clouds.

Especially in the Cygnus region the radial velocities show a pronounced group motion, which is responsible for one wave in the double period of the radial velocities. In the opposite region of the sky there are among the more distant stars no members of the Cygnus family and the result will be the appearance of a double period.

Oort[4] believes that the second harmonic term applies to the stars in general and that it can be explained on the basis of Lindblad's and his theory of galactic rotation whereby the stars are supposed to move in nearly circular orbits around the center of the galactic system. The behavior of the stars belonging to the classes just mentioned is suggestive of such an hypothesis, but if we remember that a force acting proportional to the inverse square law must produce a shear in every star cloud, it is obvious that after a relatively short time, all structure in the galaxy would be smoothed out. We must remember that just the stars that show the presence of a second harmonic term are those that are mostly clustered into nests. Computation shows that the more extensive star clouds cannot stand even one revolution of the galactic system under the influence of this shearing force. The hypothetical central mass producing this force is still a *deus ex machina*. On the other hand, the fact that the computed direction of the center of revolution, $l_0 = 325°$, coincides closely with the center

of the system of globular clusters has been emphasized as supporting the theory very strongly. However, as I have shown elsewhere[5] the application of the formula to non-B stars indicates the apparent center to be in the direction of 351°. A difference of 26° for a second harmonic term is a serious matter.

The constant of rotation derived from proper motions appears to be several times larger than anticipated by the radial velocity results. Whereas the amplitude of the second harmonic term according to Lindblad and Oort's theory should increase proportionally to the distance of the stars considered (1), this amplitude in angular measure for the proper motions should be independent of distance. I have shown in an indirect way that this is not the case.[6]

The recent observations of faint B3–B5 stars by Plaskett do not show a second harmonic term, if proper allowance is made for the solar motion. In view of the very small mean error, this negative result has a high weight, and is against the rotation theory. I do not believe that Lindblad and Oort's theory holds. The presence of a second harmonic term in some cases and its absence in other is, in my opinion, a consequence of a mixture of stars belonging to different streams.

The galactic system is built up of a number of extensive star clouds and smaller clusters, which may or may not be nuclei belonging to the larger clouds. Our sun occupies a place between the Cygnus and Monoceros clouds. In the Cygnus cloud the bulk of the B stars belong to a different stream from those in the Orion region, and the stars near the sun may, so far as their motions are concerned, be affiliated with one or the other. The presence of more clouds complicates the matter but the analysis of the motions is only a matter of time. In the case of the B stars, which are strongly clustered, the distribution in the sky affords the natural starting point in finding out the group motions, but for the stars in general, statistical methods have to be applied.

[1] *Science*, **71**, 225 (1930).

[2] *Astrophysical Journal*, **61**, 363 (1925).

[3] *Astronomische Nachrichten*, **218**, 5229–30 (1922). A recent revised theory by the same authors appeared in *Publikationen des Astrophysikalischen Observatoriums zu Potsdam*, **26**, 3 (1928).

[4] *Bulletin Astronom. Inst. Netherlands*, **3**, 275 (1927); **4**, 79, 91, 93 and 269 (1928).

[5] *Astronom. J.*, **39**, 17 (1928).

[6] *Ibid.*, **39**, 143 (1929).

GENETICAL AND CYTOLOGICAL STUDIES OF SEMISTERILITY AND RELATED PHENOMENA IN MAIZE†

BY C. R. BURNHAM*

BUSSEY INSTITUTION, HARVARD UNIVERSITY

Communicated March 12, 1930

This paper is a preliminary report of studies on two new partially sterile lines of maize derived from a pedigreed strain of genetic material grown at the University of Wisconsin in 1927. The parent line is unrelated to the stock in which semisterile-1, reported by Brink,[4] arose. In semisterile-1, approximately 50 per cent of the pollen and ovules abort and, when selfed or crossed with normals, the plants throw a ratio of 1 normal : 1 semisterile. In order to account for these results and for the origin of semisterility, Brink[4] and Brink and Burnham[6] advanced the hypothesis that a chromosomal change involving non-homologues had occurred. Simple translocation of a portion of one chromosome to a non-homologue would give a plant in which, with random distribution in the two pairs, 50 per cent of the spores either would be deficient or would possess a portion in duplicate. This unbalance was assumed to be sufficient in both cases to cause abortion. From a semi-sterile plant selfed, a new normal type was isolated which, when crossed with standard normals, gives all semisterile plants in F_1.[6] The new normal, which should be homozygous for the translocation, was designated *x*-normal, as contrasted with the standard or *o*-normal. It has been found also that semisterile-1 shows linkage with factors in two different groups, *P br* and *B lg*.[5]

To explain somewhat similar phenomena in Datura, Belling and Blakeslee[2] advanced the hypothesis of segmental interchange between two non-homologous chromosomes. Semisterility in Stizolobium, originally explained as due to two factors, has been reinterpreted more recently by Belling[1] on the basis of such interchange. According to this hypothesis, with random distribution, 50 per cent of the spores are deficient in chromatin material and are assumed to abort. A diagram of the assumed chromosomal constitution is given in figure 1. The combinations A_1 B_2 and A_2 B_1 are assumed to abort, while A_1 B_1 and A_2 B_2 form apparently normal spores. A plant with the constitution A_2 A_2 B_2 B_2 would correspond to the *x*-normal type in maize. Both the interchange hypothesis and the simple translocation hypothesis fit the breeding data thus far obtained in maize. Although duplication of chromatin material may cause abortion in certain cases, as in the simple trisomic Datura called Echinus where about half of the pollen is devoid of starch,[3] deficiency seems a more probable cause of abortion. This favors the idea of segmental interchange.

In maize it should be possible to obtain critical genetic evidence as to the type of change in different semisterile lines.

Classification for different degrees of sterility is based on estimates of the relative amounts of normal and aborted pollen. The two classes are easily distinguishable. The aborted grains are either entirely devoid of starch or nearly so, while the normal ones are well filled. Counts for sterility on the ear are made most easily from two to three weeks after pollination. For cytological study of meiosis in microsporocytes, the aceto-carmine smear method was used.

At diakinesis of the first meiotic division in normal plants ten bivalent chromosomes are found regularly. At the corresponding stage in semisterile-1 only eight bivalents occur, plus a group of four chromosomes arranged either in a ring or in a chain. In occasional cells there are ten bivalents. Apparently some chromosomal change has occurred which causes the association of non-homologues in this semisterile race. With the conjugation of homologous ends, ring configuration might be expected if segmental interchange had occurred; although, with crossing-over in the "four-strand" stage, rings might occur as a result of simple translocation. Referring to figure 1, which represents segmental interchange, the order of the chromosomes in the ring would be A_1 A_2 B_1 B_2. Only when alternate chromosomes in the ring go to the same pole would normal spores result. When adjacent non-homologues go to the same pole, i.e., A_2 with B_1 and A_1 with B_2, aborted spores result. More study is necessary to determine the detailed behavior of the altered chromosomes. The longest chromosome pair appears to be included in the group of four. Since semisterile-1 is linked with the *B lg* and *P br* groups, and since *B lg* is the third from the largest chromosome,[10] the longest one in maize must carry the factors in the *P br* group.

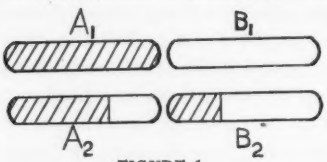


FIGURE 1

Diagrammatic representation of the assumed chromosomal constitution of a semisterile plant on the basis of segmental interchange. One end of chromosome A has interchanged with an end of chromosome B.

The new partially sterile lines were derived from sibs in a progeny of twenty-six plants from a homozygous waxy ear. Thirty-four plants from the same ear have been grown since, but all were normal. Crosses were obtained with only two of the partially sterile plants.

The first partially sterile line to be considered shows the typical breeding behavior of a semisterile. An average for eight plants showed 56.9 per cent pollen abortion. Although more than 50 per cent sterile, it breeds as a semisterile and will be referred to hereafter as semisterile-2. The cross of semisterile-2 with *x*-normal-1, i.e., the *x*-normal of semisterile-1, gives all partially sterile plants in F_1. About half of these are semi-

sterile. The other half are more than 75 per cent sterile, and will be referred to as the 75+ per cent sterile class. Pollen counts on eleven plants in the latter class gave an average of 78.5 per cent sterility. Kernel counts on eight of the same plants gave an average of 83.7 per cent ovule abortion. These results in F_1 indicate that semisteriles-1 and -2 are different. Had they been similar, there would have been a ratio of 1 normal : 1 semisterile. As far as tested, the semisterile F_1 plants are of the semisterile-1 type. Plants in the 75+ per cent sterile class were crossed with standard normals. The results from these crosses, given in the last line of table 1, show a close approximation to a ratio of one normal: one 75+ per cent sterile : two semisterile. These are the results expected if the two semisteriles are entirely independent of each other. The intermediate classes will be considered later.

TABLE 1

SEGREGATION FOR WAXY AND PARTIAL STERILITY IN THE F_1 FROM THE CROSS OF NORMAL WITH (75+ PER CENT STERILE FROM: SEMISTERILE-2 *wx wx* × SEMISTERILE-1 *Wx Wx*)

	NUMBER OF PLANTS									
	NORMAL		75+ PER CENT STERILE		SEMISTERILE		INTERMEDIATE 30 ± PER CENT STERILE		65 ± PER CENT STERILE	
	Wx Wx	*Wx wx*	*wx wx*	*Wx Wx*	*Wx wx*	*wx wx*	*Wx Wx*	*Wx wx*	*wx wx*	*Wx Wx*	*Wx wx*	*wx wx*
Cross A, Normal = *Wx Wx*	74		10	18		61	86		80	1	..	1	9
	Wx wx		*wx wx*	*Wx wx*		*wx wx*	*Wx wx*		*wx wx*	*Wx wx*	*wx wx*	*Wx wx*	*wx wx*
Cross B, Normal = *wx wx*	8	1		3	6		11	10		1	2	..	1
Total	82	11		21	67		97	90		2	2	1	10
Total	93			88			187			4		11	

The above conclusion is supported also by cytological evidence. At diakinesis in microsporocytes of the 75+ per cent sterile plants, there are six bivalent chromosomes plus two separate groups of four chromosomes each. Two separate groups would be expected if semisterile-2 arose as the result of an interchange or a translocation involving two non-homologous chromosomes, neither of which is involved in semisterile-1. In semisterile-2, rings are formed frequently, but chain configurations are more frequent than in semisterile-1. Possibly the size of the piece involved in the change may affect the relative frequency of chains and rings. Chains might be more frequent where the piece was very small in relation to the size of the chromosomes concerned. Two of the smaller chromosome pairs apparently are involved in semisterile-2. The deviation from 50 per cent sterility may be due to a lack of random distribution in the two pairs.

The segregation for waxy in the crosses of 75+ per cent sterile with standard normals is shown also in table 1. Linkage of semisterile-2 with *wx* should give, in cross A with *Wx Wx*, an excess of the two parental

types: normal *Wx Wx*, and 75+ per cent *Wx wx*. In cross B with *wx wx*, the reverse should be true. A 1 : 1 ratio is expected in the semisterile class with or without linkage. Although the numbers are small, apparently waxy is linked with semisterile-2, the calculated cross-over value being 17.7 per cent. The amount of recombination in the two classes is very different, but possibly this is not significant.

The plants showing intermediate degrees of sterility (table 1) have not been tested thoroughly. Cytological evidence and abnormal genetic ratios thus far obtained indicate that they are 21-chromosome plants. I am indebted to Dr. McClintock for suggesting the hypothesis that these intermediate types have an extra chromosome of one of the kinds composing the ring. The presence of an extra chromosome in a spore enables certain deficient combinations to survive which otherwise would abort, thus markedly reducing the amount of sterility. The four plants belonging to the class intermediate between normal and 50 per cent sterile showed 27.1, 32.9, 33.1 and 39.8 per cent pollen sterility, respectively. Only two of these were examined cytologically. Both were found to have 21 chromosomes. At diakinesis in the one showing 32.9 per cent aborted pollen there is usually a group of five chromosomes in addition to eight bivalents, making a total of twenty-one. In certain cells, the group forms a chain; in others a single chromosome is attached at its ends to two bivalents, showing that the extra chromosome belongs to one of the groups involved in semi-sterility. If, in partially sterile 20-chromosome plants, three chromosomes of the quadripartite ring go to the same pole, spores with $n + 1$ chromosomes will be formed. Anaphases have been counted in which there were eleven chromosomes at one pole and only nine at the other. The amount of sterility in the resulting 21-chromosome plant depends on which three chromosomes in the ring went to the same pole to form the $n + 1$ spore. Referring to figure 1, the two types of combinations possible in the $n + 1$ spore would be A_1 A_2 with B_2 and A_1 A_2 with B_1. The amount of sterility also depends on whether or not there is random pairing in the set containing the extra chromosome. There might be preferential pairing of the two normal chromosomes. The frequency with which the extra chromosome lags at meiosis and is eliminated also will affect the degree of sterility.

From 75+ per cent sterile plants, there should be two classes of intermediates: those falling between semisterile and normal, and those between semisterile and 75+ per cent sterile. The former class is distinct; but in the latter there is probably some overlapping with 20-chromosome semisterile-2 plants which show an average of about 57 per cent abortion. The upper class of intermediates has not been examined cytologically.

Data from two 21-chromosome plants backcrossed to a *c sh wx* stock are given in table 2. On the basis of interchange as represented in figure 1,

both plants are assumed to have arisen from an $n + 1$ egg of the type A_1 A_2 B_2. The extra chromosome in B11–13 belongs probably to the *sh wx* set. In the crosses with both plants, the ratios for *sh* and *c* deviate widely from the normal 1 : 1 ratios obtained from semisterile sibs. Close fits to the observed numbers and also to the degree of sterility may be obtained by assuming the proper amount of preferential pairing and the proper frequency of lagging univalents in the trisomic groups. The assumption that the break in the *sh wx* chromosome occurred on the side of *wx* away from *sh*, placing *sh* about 38 units from the break, fits the observed data. Further study is necessary to determine the amounts of preferential pairing and lagging, and to determine the frequency of crossing over between *sh* and semisterile-2.

TABLE 2

SUMMARY OF BACKCROSS DATA FROM TWO EXCEPTIONAL 21-CHROMOSOME PLANTS IN THE PROGENY OF THE CROSS: [75+ PER CENT STERILE FROM (*C Sh wx* × ? *Sh Wx*)] × NORMAL *c sh wx*

CROSS	PER CENT POLLEN ABORTION	*C*				*c*				TOTAL
		Sh		*sh*		*Sh*		*sh*		
		Wx	*wx*	*Wx*	*wx*	*Wx*	*wx*	*Wx*	*wx*	
B11–13 × *c sh wx*	32.9	94	73	..	2	1	..	9	38	217
c sh wx × B11–2	27.1	..	189	..	11	..	1	..	365	566

From B11–2, the *Sh* : *sh* ratio through the pollen is approximately 1 : 2. This might appear to be a normal trisomic ratio, since in maize $n + 1$ pollen functions only occasionally.[10] However, the progeny in which plant B11–2 was found came from the cross: 75+ per cent sterile *Sh Sh* × normal *sh sh*. Since the female parent in which the $n + 1$ spore probably arose was *Sh Sh*, *Sh* seeds would have been in excess in the backcross given in table 2 if the extra chromosome were the one bearing this factor. The extra chromosome is proably the other one involved in semisterile-2. With random pairing in the trisome in a plant of the type assumed for B11–2 and B11–13, it would be possible to get apparent trisomic ratios for factors which were not in the trisome, provided they were closely linked with semisterility. In this case, where *sh* is assumed not to be closely linked with sterility, the assumption of preferential pairing in the trisome gives a close fit to the observed numbers and also to the degree of sterility.

Plant B11–2 was crossed also on two *Pr pr* plants, giving 233 *Pr*:156 *pr* seeds. The deviation from the expected 3 : 1 ratio is over ten times its probable error, and suggests that the *pr* v_2 chromosome is the other one involved in semisterile-2. This is not certain, however, since ratios for *pr* often deviate widely from the expected (Fraser, unpublished data from Cornell).

In two other exceptional plants, abnormal ratios for *waxy* were obtained in the pollen. One plant which showed 37.7 per cent aborted pollen had only 11.9 per cent waxy grains. Another plant showing 9.7 per cent

pollen sterility had 16.3 per cent waxy grains. In semisterile plants heterozygous for waxy, 50 per cent of the pollen is waxy.

The abnormal shrunken and waxy ratios furnish additional evidence that the *sh wx* chromosome is one of those involved in semisterile-2.

A partially sterile sib in the culture in which semisterile-2 was found was crossed with *x*-normal-1. All but one of the forty F_1 plants were partially sterile. The exception was a normal which may have resulted from a stray pollination. About half of the F_1 plants were semisterile, while the other half were somewhat less than 75 per cent sterile (referred to as 75− per cent). Pollen counts on nine plants in the latter class showed an average of 68.7 per cent sterility, while kernel counts on the same plants gave an average of 71.3 per cent abortion. These results indicate that the new sterile is different from semisterile-1. Since the new class from the corresponding cross with semisterile-2 showed 78.5

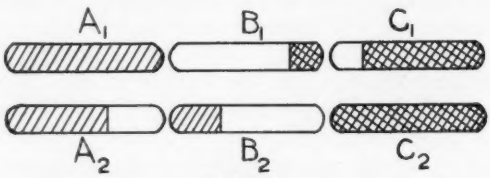


FIGURE 2

Diagram of the assumed chromosomal constitution of a 75 per cent sterile plant from the cross of semisterile-1 with semisterile-3, according to the segmental interchange hypothesis. One member of the B pair has interchanged with an A chromosome while the other member of the B pair has interchanged with a C chromosome. If homologous ends associate, a ring of six chromosomes might be expected at diakinesis.

per cent pollen abortion, the new type is also different, apparently, from semisterile-2. Plants belonging to the 75− per cent sterile class were crossed with standard normals. Out of 97 plants from these crosses, 90 were semisterile, one was normal, one was 81.5 per cent sterile and five belonged to a distinct class intermediate between normal and 50 per cent sterile. Practically every functional spore from the 75− per cent sterile plants must have been a semisterile producer. This suggests that the new type, semisterile-3, also involves interchange or translocation between non-homologous chromosomes, but that one of them is the same as is involved in semisterile-1. A diagram of the assumed chromosomal constitution of a 75− per cent sterile plant, based on segmental interchange is given in figure 2. In the pair common to both steriles, the B_1 member has interchanged with chromosome C_1, while the B_2 member has interchanged with A_2. Separation at meiosis of the members of this com-

mon pair would give only two kinds of normal spores, A_1 B_1 C_1 and A_2 B_2 C_2. In crosses with standard normals, the first would give semisterile-3 and the other would give semisterile-1.

On the above hypothesis, configurations involving six chromosomes should be found at meiosis in 75− per cent sterile plants. Cytological examination of microsporocytes at diakinesis does show a group of six chromosomes plus seven bivalents. The group is either a large closed ring or, occasionally, an open one. On the basis of Belling's interchange hypothesis illustrated in figure 2, the order of the chromosomes in the ring of six would be A_1 A_2 B_1 C_2 C_1 B_2. Only the combinations of alternate chromosomes in the group would give normal spores. The deviation from 75 per cent sterility is probably due to lack of independent distribution of the three pairs involved. Semisterile-3 may be considerably less than 50 per cent sterile; or the association of the three pairs in a ring or chain may prevent their independent distribution.

If opposite ends of the two members of the common chromosome pair are involved in semisteriles-1 and -3, as illustrated in figure 2, cross-overs between them might occur. In crosses with standard normals, these would give normal and 75± per cent sterile plants. One normal plant and one with 81.5 per cent sterility were found. They are being tested to determine whether they are the result of crossing-over or stray pollinations.

Two of the three pairs in the ring of six chromosomes must be *B lg* and *P br*, respectively, since these pairs are involved in semisterile-1. Either the *B lg* or the *P br* chromosome, therefore, must be concerned in semisterile-3. Certain configurations at diakinesis indicate that *B lg* is the one involved. Genetic tests are in progress.

The five plants belonging to the class intermediate between normal and 50 per cent sterile have been examined cytologically using either root tip or microsporocyte material. All proved to have 21 chromosomes. Where microsporocyte material was available, groups of five chromosomes were found in addition to eight bivalents, showing that the extra chromosome belongs to the group involved in semisterility. The frequency of these plants is very high, over six per cent having been found following crosses where the 75− per cent sterile plant was the female parent. The 20-chromosome, partially sterile plants thus furnish another source for the isolation of certain simple trisomics in maize.

In Pisum, an unexpected linkage of two factors which are ordinarily independent has been found by Hammarlund.[9] Håkansson[8] examined plants from this cross cytologically and found a ring of four chromosomes at meiosis. No mention of sterility was made in these papers; but by comparing the number of plants per progeny from the normal line with that from the cross which shows linkage, I find that the latter progenies

are about half as large. In the normal line the average for seven progenies is 115.6. In the cross showing linkage, the average for 28 progenies is only 53.0. This suggests that Hammarlund's double dominant line *K*, which shows linkage of the two factors in its crosses, may be an *x*-normal type. Since only the parental combinations of chromosomes in the two pairs involved in semisterility are viable, two factors closely linked with sterility but in different chromosome pairs would show apparent linkage with each other. This would be true whether or not either or both were in the interchanged or translocated piece. Unless both were linked with sterility to about the same degree, the recombination classes should not be of equal size. Examination of the data shows that the two classes are approximately the same size. Another case of ring-formation in Pisum was reported recently by Miss Richardson[12] in a cross showing 50 per cent sterility.

Ring formation in Oenothera has been interpreted by Darlington[7] on the basis of segmental interchange between non-homologous chromosomes. One of the cases described in maize in the present paper shows how configurations involving more than two pairs of chromosomes may be built up by combining the proper semisterile types.

Müntzing[11] has studied partial sterility in crosses of Galeopsis species. Although the complete story has not been published, the breeding behavior suggests a situation similar to that in maize and in Datura. Cytological examination revealed no abnormalities, but only one diakinesis preparation was examined.

Summary.—1. In semisterile-1 and in two new types, semisteriles-2 and -3, there is an association of non-homologous chromosomes at meiosis. A group of four chromosomes occurs plus eight bivalents.

2. Sufficient evidence is not available to decide definitely whether this is the result of segmental interchange or simple translocation.

3. The combination of semisteriles-1 and -2 gives a new class which is somewhat more than 75 per cent sterile, and in which two separate groups of four chromosomes occur plus six bivalents.

4. Neither of the two pairs involved in semisterile-2 is concerned in semisterile-1. The *sh wx* pair is one of those involved in semisterile-2.

5. The combination of semisteriles-1 and -3 gives a new class which is a little less than 75 per cent sterile, and in which there is a group of six chromosomes plus seven bivalents.

6. One of the pairs involved in semisterile-3 is probably the same as is concerned in semisterile-1, and, therefore, must be either *P br* or *B lg*.

7. Exceptional 21-chromosome plants showing intermediate degrees of sterility have arisen from partially sterile 20-chromosome plants. The extra chromosome belongs to the group involved in semisterility.

† This work was begun in the Department of Genetics at the University of Wisconsin.

The greater part of the analysis was made at Cornell University, certain parts being finished at the Bussey Institution.

I am indebted to Dr. R. A. Brink for giving me the semisterile stocks, to Dr. R. A. Emerson for making it possible to grow the material at Cornell University in 1929, and to Drs. L. W. Sharp and R. A. Brink for helpful suggestions in the preparation of this manuscript. To Dr. Barbara McClintock of Cornell University I wish to express my sincere thanks for her aid and encouragement in attacking the cytological aspects of the problem.

* NATIONAL RESEARCH COUNCIL FELLOW at Cornell University and Bussey Institution, Harvard University.

[1] Belling, J., *Zeit. indukt. Abstamm. Vererb.*, **39**, 286–288 (1925).
[2] Belling, J., and A. F. Blakeslee, *Proc. Nat. Acad. Sci.*, **12**, 7–11 (1926).
[3] Blakeslee, A. F., *Ibid.*, **10**, 109–116 (1924).
[4] Brink, R. A., *J. Hered.*, **18**, 266–270 (1927).
[5] Brink, R. A., (Abstract) *Anat. Rec.*, **44**, 280 (1929).
[6] Brink, R. A., and C. R. Burnham, *Amer. Nat.*, **63**, 301–316 (1929).
[7] Darlington, C. D., *J. Gen.*, **20**, 345–363 (1928–29).
[8] Håkansson, A., *Hereditas*, **12**, 1–10 (1929).
[9] Hammarlund, C., *Ibid.*, **4**, 235–238 (1923); **12**, 210–216 (1929).
[10] McClintock, B., and H. E. Hill, (Abstract) *Anat. Rec.*, **44**, 291 (1929).
[11] Müntzing, A., *Hereditas*, **12**, 297–319 (1929).
[12] Richardson, Eva, *Nature*, **124**, No. 3128 (1929).

EVIDENCE THAT NATURAL RADIOACTIVITY IS INADEQUATE TO EXPLAIN THE FREQUENCY OF "NATURAL" MUTATIONS

BY H. J. MULLER AND L. M. MOTT-SMITH

DEPARTMENT OF ZOÖLOGY, UNIVERSITY OF TEXAS, AND DEPARTMENT OF PHYSICS, THE RICE INSTITUTE

Communicated February 28, 1930

I. The Problem.—When it was found that irradiation of Drosophila by artificial means results in a pronounced increase in the mutation frequency, the question immediately arose (Muller, 1927) whether the mutations that occur in untreated material are caused by radiation of a similar type from radioactive substances naturally occurring. Since this problem was raised, results have been published by Olson and Lewis, Babcock and Collins and others, which have been taken as indicating that such was actually the case.

II. The Inadequacy of the High-Energy Radiation from the Outer Environment.—In order to test the above possibility in our material, we have made calculations of the relation between the artificial and natural radiation effects, for comparison with the artificial and "natural" mutation frequencies. As measures of the intensity of the radiation in both cases, we have taken the ionization per cc. per second in air. In order, then,

to compare the total effect of each type of radiation, this intensity factor must be multiplied by the duration of exposure. An empirical test of the validity of using the ionization produced in air as a measure of the intensity of high-energy radiation here effective has been furnished by Hanson's experiments with radium and x-rays in which he found that the induced mutation frequency was always exactly proportional to the ionization in air, regardless of the type of rays used.

We were enabled to gain fairly definite information concerning the intensity of the artificial radiation by means of a "dosimeter"* which gives the intensity of the x-rays directly in "*r*" units per minute. With a treatment commonly used by us (termed "*t* 12, new machine," in previous publications) we found an intensity of 81.4 *r* units per minute. This treatment lasted 42 minutes and therefore involved 3420 *r* units. One *r* unit represents the production of one electrostatic unit of charge per cc. in air, which corresponds to 2.1×10^9 (pairs of) ions per cc. The total number of ions per cc. produced in air by the entire artificial treatment is therefore $3420 \times 2.1 \times 10^9$, or 7.2×10^{12}, for the 42 minutes. This figure is probably somewhat too small, owing to the fact that the dosimeter may not be affected sufficiently by the softer radiation, but it will be seen below to be of major importance, for our comparison, that the figure arrived at for our artificial treatment should not err on the side of being too large.

As a check on the value found by means of the dosimeter, the ionization produced by the x-radiation was also computed by an independent method. The intensity of the x-radiation emitted by a tungsten-target Coolidge tube such as we used is known as a function of the power input. Using the data of Terrill, approximate calculations of the intensity under our conditions of distance (13.5 cm.), screening (one mm. of aluminum), voltage (50 k. v. peak) and current (10 m. a.) could readily be made. Having computed the approximate intensity, 1.2×10^{-4} Watts per sq. cm., and effective wave-length (0.52 Å), and taken into consideration the absorption coefficient of the medium (air), the number of photo-electrons emitted per cc. per second was obtained. This number was 1.3×10^7. By assuming that each photo-electron produces 250 ions, the total number of ions produced per cc. per second was obtained. This value, 3.2×10^9, when multiplied by the number of seconds (42×60), gives the total number of ions per cc. produced by the treatment as 8.2×10^{12}. The agreement between the number so obtained and the value, 7.2×10^{12}, arrived at by the dosimeter measurement, is satisfactory. Although the second method, because of some uncertainty regarding the exact average wave-length, tube activity, etc., must be considered as giving only an approximate verification, the agreement found certainly indicates that there is no gross error.

The amount of ionization in air due to radioactivity occurring in the outer environment has been studied by many observers. It is found to vary considerably for different localities, different distances above the earth's surface, etc. A study of this evidence shows that the ionization from these causes in the vicinity of the flies is very unlikely to be greater than 30 ions per cc. per second. The flies are subjected to this radiation throughout their entire reproductive generation (from germ cell of parent to germ cell of offspring). This has ordinarily been about two weeks. The total number of ions per cc. produced in air by the radiation to which they are subject is then not more than thirty times 1.2×10^6 (the number of seconds in two weeks), or 3.6×10^7.

Artificial irradiation with the dose of 3420 *r* units results in about 15 recessive lethal mutations per hundred treated X-chromosomes. This value has been repeatedly and consistently obtained at the Texas laboratory by Muller, by Harris and, with greater accuracy, by Oliver, when spermatozoa were treated, the fluctuations in mutation frequency being only as great as would be expected in view of the fluctuations of the x-ray machine and the error due to random sampling. It should be stated, however, that this value probably represents an upper limit of mutability for the given dose of x-rays, since experiments with cells other than sperm cells—viz., spermatogonia, oögonia and oöcytes of the adult, and germ cells of the larva—indicate that the induced mutation frequency in them is only one-tenth to one-half as high as the above. In order to stay "on the side of caution," however, we will adopt this high mutation frequency of the treated sperm cells as representing the value regularly to be obtained from artificial treatment.

In untreated material the frequency of mutations of the kind in question is subject to considerable variation, the values having been observed by Muller and Altenburg to range from one or more per hundred down to about one per thousand. Recently Hanson, and Babcock and Collins have reported mutation rates in untreated material, and these rates have, in general, agreed in falling between the limits just stated. The lowest "natural" value then is about 1:150 of the highest value artificially induced by the 3420 *r* dose.

It has been shown by Hanson and by Oliver that the artificially induced mutation frequency is strictly proportional to the amount of ionization produced in air, calculated as above. If, then, the natural mutations are produced by natural radiation, we should find the amount of ionization occurring without special treatment to be 1:150 of the amount produced artificially. Comparing the value, 7.2×10^{12} ions per cc. for the artificial treatment, with that of 3.6×10^7 ions per cc. for the untreated material, we find a ratio of 1:200,000. The difference between these two ratios, 1:150 and 1:200,000, is undoubtedly significant, particularly since we have

taken the values most unfavorable to our case, namely, the minimum possible value for the natural mutation frequency, the maximum value for the artificially induced mutation frequency, the maximum for the natural ionization and the minimum for the artificially induced ionization. The fact that the former ratio (1:150) is nevertheless about thirteen hundred times the latter ratio (1:200,000) shows that the natural mutation frequency is at least thirteen hundred times as high as it would be if it were caused solely by the radiation which the flies receive from their outer environment.

III. The Inadequacy of Radiation from Weakly Radioactive Substances in or near the Organisms.—We have used the term "outer environment" in a rather specialized sense, not including the medium immediately in contact with the flies (their food), or any substances within the organisms by which the germ plasm may be affected. The possibility remains to be examined that these may contain significant amounts of radioactive material.

There is, for example, an appreciable amount of potassium in the protoplasm and body fluids, as well as in the food. It is known that the ionization produced in air by potassium (which gives off beta rays) is such that one gram of it is capable of yielding a current of about 0.0014 electrostatic units, or three million ions per second. Now the maximum concentration of potassium which might be found in the organism or its food would be about half a gram per hundred cc. (i.e., 1:200). This is, in fact, far more than ordinarily occurs in animal tissues. Potassium, if distributed uniformly in this concentration in air, would in two weeks (1.2×10^6 seconds) yield $1:200 \times 3{,}000{,}000 \times 1.2 \times 10^6$, or 1.8×10^{10} ions per cc.

To compare the potassium effect in the organism with the effect of the artificial radiation, we cannot directly use the values for ionization in air. This is because the effect of weakly penetrating radiation produced *in situ* in a medium (as the radiation from potassium is) does not vary with a change of the medium as does the effect of radiation entering the medium from outside sources. The potassium may be considered to be distributed throughout a considerable volume of the very space in the interior of which the measurements are to be taken, and most of the energy from its radiation must be absorbed in this space regardless of whether the medium be protoplasm or air. Thus the figure of 1.8×10^{10} ions per cc. for potassium in air may be taken as substantially representing the maximum "effectiveness" of the radiation from the potassium in the protoplasm also (although in the protoplasm not as much of the energy absorbed may actually be expended in the specific form of ionization). On the other hand, the energy absorbed from the artificial radiation (and from the greater proportion of the natural radiation from outside sources) varies with the coefficient of absorption of the medium. The absorption, when

air and protoplasm are compared, is nearly proportional to their respective densities. Hence we must multiply the figure 7.2×10^{12} (the number of ions per cc. produced in air by the artificial radiation) by 800, representing the ratio of the densities of protoplasm to air. This gives 5.8×10^{15} as the figure representing the "effectiveness" (in the sense used above) of the artificial radiation in the protoplasm. It will be seen that the ratio of the "effectiveness" of the potassium to that of the artificial radiation is $\frac{1.8 \times 10^{10}}{5.8 \times 10^{15}}$, or $\frac{1}{3.2 \times 10^{5}}$, which is only $\frac{1}{2100}$ of the ratio (1:150) of the natural to the artificial mutation frequencies. The maximum radiation from the potassium then is even less adequate than that from the outer environment to account for the observed natural mutation rate.

The ionization produced by uranium (chiefly by its alpha rays) is of the order of a thousand times that produced by an equal weight of potassium. Therefore $^1/_2$ gram of uranium per 100 cc. would be 1000 times as effective as $^1/_2$ gram of potassium per 100 cc., and hence this concentration of uranium would be of the right order to give the observed natural mutation frequency. Such a concentration of uranium as this is, however, practically out of the question in the case of a living organism. Similar considerations apply to thorium. Direct measurements are, however, called for to determine whether more highly radioactive materials are present in sufficient amounts to explain the mutation frequency.

IV. The Paucity of Radiation from Other Sources.—Measurements were accordingly undertaken of the total radioactivity of the flies and of their food (standard banana-karo-agar). About two and a half grams of recently killed *Drosophila melanogaster* (approximately 5000 in number, comprising various stocks mixed together) were desiccated in an oven at 120°C. for three days and then ground to a powder. This powder was spread over an area of 36 sq. cm. in a standard type of ionization chamber. The ionization current, produced by an accelerating potential of 270 volts, was measured by means of a quadrant electrometer.

In order to determine the sensitivity of the apparatus, the ionization chamber was calibrated by measurements of the activity both of a potassium chloride surface and also of a small uranium standard. The current obtained from the empty ionization chamber was also measured directly and was found to be 2.5×10^{-15} amperes, representing an effect due to "natural" radiation. The probable error of this determination (based on repeated measurements) was found to be such that a change in this current of about 9% would be three times the probable error of the difference and hence would have been detectable. The smallest detectable activity then corresponds to a current of 9% of 2.5×10^{-15} amperes, or 2.25×10^{-16} amperes. Since the area was 36 sq. cm., this corresponds to an activity per sq. cm. of 6.25×10^{-18} amperes. It is known that one

sq. cm. surface of uranium oxide (U_3O_8) gives 5.8×10^{-13} amperes. Comparing these two figures, and allowing for the fact that the uranium oxide is itself only 85% uranium, we see that in our experiment we could have detected a concentration of uranium of $\frac{6.25 \times 10^{-13}}{5.8 \times 10^{-13}} \times 0.85$, or about one part uranium in 100,000 parts of the dried fly material or food. An equal amount of activity on the part of other radioactive substances would have been similarly detectable, of course. The direct measurements on the current produced in the presence of the uranium and the potassium salts were in substantial agreement with this determination of the sensitivity of the apparatus.

Numerous measurements were made of the ionization in the presence of the fly powder, and no evidence of any effect of the material on the ionization current was found. Similarly negative results were obtained in tests of the dried and ground fly food. We can, therefore, conclude that the tested materials had a radioactivity corresponding to less than one part in 100,000 of uranium. It will be recalled that about one part of uranium in 200 parts of the fresh fly material, or one in 50 parts of the dry fly material, would have been necessary to account for the mutation frequency. This required amount is 2000 times the amount we could have detected; hence there can be no doubt that any radioactive material present was entirely inadequate in quantity.

An article has recently been published by Vernadsky† in which he reports finding that in various organisms, both plants and animals, a considerably higher concentration of radium is present than in the water with which they come into contact, and he consequently regards organisms as collectors of radium. The proportions of radium which he finds are nevertheless very low, of the order of 1:50 of the lowest values that we could have detected. They are, therefore, quite negligible for our present purposes, even if it be supposed that other radioactive substances are stored similarly to the radium.

If the "natural" radioactivity from all terrestrial (including internal) sources combined is thus inadequate, there is no possibility that the cosmic radiation could play a significant role either, since the ionization produced by it (even at high altitudes) is considerably less than that produced over the land by the radiation from terrestrial sources.

V. Consideration of a Possible Alternative.—The only escape remaining from the conclusion that causes other than high-energy radiation are responsible for the production of the great majority of the mutations in untreated *Drosophila* would now lie in the assumption that the germ cells or tissues immediately surrounding them are specific absorbers of radioactive material, and store it in a concentration exceeding by over a thousand times the concentration present in the organism as a whole. Such

tissue would have to form a very small proportion of the fly since even if its weight formed as little as 1:1000 of the total, the consequent rise in radioactivity of the whole mass would have been detectable. We find by measurement that the male germ cells (mostly mature) form, at a very conservative estimate, over 1:100 of the total volume of the fly, and the eggs considerably more. The mature germ cells, therefore, cannot themselves be receptacles of the postulated amount of radioactive material, and it would have to be supposed that the cells in the gonad kept passing the material along, as they divided, so that it remained condensed in the immature cells.

A possible biological test of such a special hypothesis might be provided by a study of the effects of feeding the flies small measured quantities of highly radioactive material, such as would not raise the mutation frequency appreciably if uniformly distributed, but would if stored near the germ cells. The amount of increase of the concentration of radioactive substances in the flies, as a whole, so produced, could then be measured directly, and compared with the increase, if any, in the frequency of mutations. Thus it could be determined whether or not the tissue in the neighborhood of the germ cells stored the material to a much greater extent than the fly as a whole did.‡ Such tests would have to be made with various highly radioactive materials, since these differ in their chemical properties. Unfortunately, a negative result from such tests would not in itself be demonstrative, since it might be held that the concentration was automatically fixed at a certain amount, beyond which it was not allowed to rise. (Thus Vernadsky finds that organisms of each species have their own characteristic concentration of radium.) We believe, however, that the assumption of such an extremely high concentration of radioactive material as would here be necessary, confined to the immediate vicinity of the germ cells, is *per se* very implausible, especially in consideration of the fact above referred to that the mature germ cells do not carry any such concentration themselves.

VI. Conclusions.—It is accordingly probable that natural radioactivity is not the major cause of mutations, and of organic evolution, but that most mutations come about as a result of other causes. The search for these causes must continue. It would still remain true that a certain number of mutations in untreated flies must be caused by natural radiation, but the ratio of these to all the natural mutations is probably smaller than one per thousand, so that it is not likely that more than one or two, if any, among all the hundreds of mutations so far observed in the different species of Drosophila have been caused in this way. At the same time, the gene mutations resulting from irradiation are indistinguishable from those otherwise caused, and this method of artificially producing them in quantities, the only one at present known, may be

used to advantage in the study of the mechanism of mutation and of the properties of the genetic material.

* At the instance of the Committee on the Effects of Radiation on Living Organisms, of the National Research Council, this instrument was supplied at cost by the Victoreen Company. The expenses in connection with this investigation were defrayed by a grant from this committee.

† We are indebted to Dr. Timofeëf-Ressovsky for translating Vernadsky's paper for us. Dr. Timofeëf-Ressovsky informs us that he has been making an investigation similar to our own. We have also received a communication from Dr. Vernadsky, calling attention to recent measurements of Brunovsky,[3] which show that the concentrations of radium in organisms are really lower than the preliminary results indicated.

‡ Direct measurements of the radioactivity of different gonadic tissues could be made in the case of larger animals, but in the absence of knowledge of their mutation rates, naturally, and when artificially irradiated, such information would be of little value for our present purposes.

[1] Babcock, E. B., and Collins, J. L. '29a. "Does Natural Ionizing Radiation Control Rate of Mutation?" *Proc. Nat. Acad. Sci.*, **15**, 625–628.

[2] ——. '29b. "Natural Ionizing Radiation and the Rate of Mutation." *Nature*, **124**, 227–228.

[3] Brunovsky, B. K. '29. "On the Concentration of Radioactive Substances in Living Organisms." (Russian.) *Trav. d. Lab. Biogéochim près l'Acad. d. Sci. d. U. S. S. R.*, **1**, 75.

[4] Dobrovolskaia-Zavadskaia, N. '29. "The Problem of Species in View of the Origin of Some New Forms in Mice." *Biol. Rev.*, **4**, 327–351.

[5] Haldane, J. B. S. '29. "The Species Problem in the Light of Genetics." *Nature*, **124**, 514–516.

[6] Hanson, F. B., and Heys, F. '29a. "An Analysis of the Effects of the Different Rays of Radium in Producing Lethal Mutations in Drosophila." *Am. Nat.*, **63**, 201–213.

[7] ——. '29b. "Duration of the Effects of X-Rays on Male Germ Cells in Drosophila Melanogaster." *Am. Nat.*, **63**, 511–516.

[8] ——. '30. "A Possible Relation between Natural (Earth) Radiation and Gene Mutations." *Science*, **71**, 43–44.

[9] Hanson, F. B., and Stanton, E. '29. "The Effect of Increasing X-Ray Voltage on the Production of Lethal Mutations in Drosophila Melanogaster." (Abstr.) *Anat. Rec.*, **44**, 293.

[10] Harris, B. B. '29. "The Effects of Aging of X-Rayed Males upon Mutation Frequency in Drosophila." *J. Hered.*, **20**, 299–302.

[11] Joly, J., and Dixon, H. H. '29. "Cosmic Radiations and Evolution." *Nature*, **123**, 981.

[12] Muller, H. J. '27a. "Artificial Transmutation of the Gene." *Science*, **66**, 84–87.

[13] ——. '27b. "The Problem of Genic Modification." *Verh. d. V. Intern. Kong. f. Vererb. I* (*Zeit. f. ind. Abst. u. Vererb. Supl. Bd. I.*, publ. 1928), 234–260.

[14] ——. '28a. "The Production of Mutations by X-Rays." *Proc. Nat. Acad. Sci.*, **14**, 714–726.

[15] ——. 28b. "The Measurement of Gene Mutation Rate in Drosophila." *Genetics*, **13**, 279–357.

[16] ——. '29. "The Method of Evolution." *Sci. Mo.*, **29**, 481–505.

[17] ——. '30. "Radiation and Genetics," *Amer. Nat.* (in press).

[18] Oliver, C. P. '30. "The Effect of Varying the Duration of X-Ray Treatment upon the Frequency of Mutation." *Science*, **71**, 44–46.

[19] Olson, A. R., and Lewis, G. N. '28. "Natural Radioactivity and the Origin of Species." *Nature*, **121**, 673–674.

[20] Stadler, L. J. '28. "The Rate of Induced Mutation in Relation to Dormancy, Temperature, and Dosage." (Abstr.) *Anat. Rec.*, **41**, 97.

[21] Terrill, H. M. '26. "The Energy of X-Rays." *Phys. Rev.*, **28**, S II, 438–443.

[22] Timofeëf-Ressovsky, N. W. '29. "Der Stand der Erzeugung von Genovariationen durch Röntgenbestrahlung." *J. Psych. Neurol.*, **39**, 432–437.

[23] Vernadsky, V. I. '29. "On the Concentration of Radium in Living Organisms." (Russian.) *C. R. de l'Acad. d. Sci. d. U. S. S. R.*, **A2**, 33–34.

THE INFLUENCE ON THE IDEAL SOLUTION LAWS OF THE DISTRIBUTION OF POLARITY WITHIN THE MOLECULE

BY J. H. HILDEBRAND AND J. M. CARTER

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF CALIFORNIA

Communicated March 5, 1930

It has long been known that differences in the degree of polarity between two molecular species tend to produce, in their solutions, deviations from the ideal solution laws.[1] Recent progress in clarifying our ideas concerning polarity, which we owe principally to Debye,[2] has made it possible to measure the electric moments of molecules and to determine how they are influenced by the nature and arrangement of their parts, and it is now possible to test an opinion expressed earlier by the senior author[3] that to understand the effect of polarity on the solution laws it is often necessary to take account of the polarity of the parts of a molecule, and not simply the polarity of the molecule as a whole.

There are only a few systems for which all the necessary data are at hand, and the best for our purpose appear to be the solutions of benzene with nitrobenzene, with the three dinitrobenzenes and with 1-3-5-trinitrobenzene. According to Williams[4] the electric moment of benzene is very small, $<0.1 \times 10^{-18}$ e.s.u., that of nitrobenzene is 3.90×10^{-18}, for the ortho-, meta- and para-dinitrobenzenes the moments are 6.05×10^{-18}, 3.81×10^{-18} and 0.32×10^{-18}, respectively; for trinitrobenzene the moment is 1.08×10^{-18}. Solutions of nitrobenzene and benzene show deviations from Raoult's law (*vide infra*) in harmony with the difference between their electric moments; if, then, deviations from Raoult's law depend upon the electric moments of the molecule as a whole, we would expect very little deviation for solutions of benzene with para-dinitrobenzene, more for benzene with meta-dinitrobenzene, and most with benzene and ortho-dinitrobenzene. If, on the other hand, the forces between two molecules are determined chiefly by the polarity of the parts which are nearest

to each other, there should be much less difference in the behavior of the three dinitrobenzenes toward benzene.

We have measurements of solubility for the system benzene-nitrobenzene by Dahms,[5] for benzene with the three dinitrobenzenes by Kremann.[6] Since these substances have different melting points and heats of fusion we cannot directly compare their solubility curves, but we can make a comparison sufficiently accurate for our purpose by means of the equation used by the senior author[7] to define a "regular" solution,

$$\bar{F}_2 - \bar{F}_2^i = b N_1^2 \qquad (1)$$

where $\bar{F}_2$ is the partial molal free energy of component X_2 in the regular solution, $\bar{F}_2^i$, its partial molal free energy in an ideal solution, b, a constant which expresses the deviation from Raoult's law and N_1 the mole fraction of the other component, X_1. For regular solutions Eq. (1) is considered to be independent of the temperature; for the solutions here under consideration this may not to be strictly true, but it may, nevertheless, be regarded as sufficiently approximate to serve for the rather rough comparison we wish to make, and the data themselves indicate that this is the case. The free energy of transfer of a mole of X_2 from the ideal to the actual solution is given by

$$\bar{F}_2 - \bar{F}_2^i = RT \ln(N_2^i/N_2); \qquad (2)$$

the ideal solubility, N_2^i, is calculated from the heat of fusion, ΔH_f, and the absolute melting point, T_m, by the familiar equation:

$$\log N_2^i = \frac{-\Delta H_f}{2.3R}\left(\frac{1}{T} - \frac{1}{T_m}\right). \qquad (3)$$

Values for the heats of fusion and melting points of the several nitrobenzenes were taken from Andrews, Lynn and Johnston.[8] It is hardly necessary to tabulate the figures used and the results of the calculations, since the conclusion is sufficiently well indicated in figure 1 where $\bar{F}_2 - \bar{F}_2^i$ is plotted against N_1^2. The number and positions of the nitro-groups have been indicated in a manner that is self-explanatory.

It will be seen that a single straight line serves for all three dinitrobenzenes, which shows not only that their behavior is satisfactorily expressed by Eq. (1) but that *there is no significant difference between their behaviors toward benzene in spite of the great differences in their electric moments.*

The point for nitrobenzene represents the eutectic point and is doubtless sufficient to give the slope with satisfactory accuracy even though higher values of N_1 are thereby excluded. It is a very striking fact that the slope

of this line is approximately half the slope of the line for the dinitrobenzenes. It will be remembered that this slope represents b in Eq. (1) and expresses the deviation of the solution from the ideal. We may say, accordingly, that the substitution of the nitro-group in the benzene ring causes a deviation from Raoult's law expressed by $b = 320$ cals., while the substitution of two nitro-groups gives rise to nearly twice this deviation, making $b = 580$ cals. One must not, of course, attach too great significance to the apparent simplicity of this relationship unless it is later confirmed for other systems. It is to be hoped that data will be obtained with this end in view.

There are data obtained by Desvergnes[9] for the solubility in benzene of symmetrical trinitrobenzene. Unfortunately, we have no figure for its heat of fusion, but since its solubilities[10] in ortho-dinitrobenzene, in

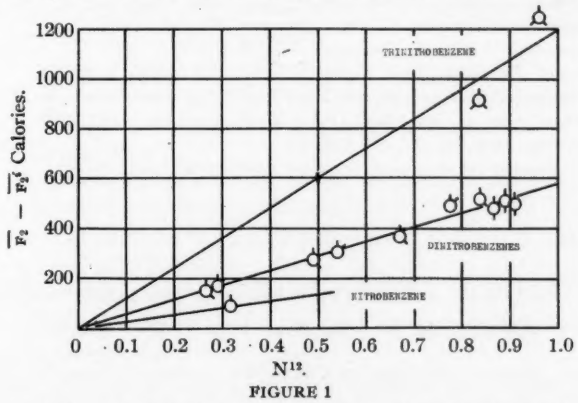


FIGURE 1

Deviations from Raoult's Law shown by solutions of benzene with various nitrobenzenes.

meta-dinitrobenzene and in 2-4-6-dinitrotoluene are all practically identical they may safely be assumed to be approximately ideal. Fixing the ideal solubility in this way the solubilities in benzene show a departure from Raoult's law, indicated in figure 1, which is approximately twice that for the dinitrobenzenes, corresponding to $b = 1200$ cals. as compared with 580 cals. for the dinitrobenzenes and 320 cals. for nitrobenzene. The addition of the third nitro-group in the symmetrical arrangement apparently leaves so small a non-polar portion exposed that the unlikeness to benzene is more than proportionately increased, although the electric moment of the molecule is thereby much reduced.

We may mention, further, that the three dinitrobenzenes, in spite of the differences in their electric moments, show mutual solubilities agreeing closely with Raoult's law.

The combined weight of this evidence indicates that *it is the number and polarity of the substituent groups rather than the electric moment of the whole molecule which determines deviations from Raoult's law.*

We must not overlook, however, the fact that when the field of a polar bond is sufficiently buried within the molecule its influence upon solubility tends to disappear. Thus, in spite of the polarity of the bond between the carbon atom and the nitro-group, as shown in nitrobenzene and nitromethane, we might expect tetranitromethane to behave in its solutions as a substance of low polarity. Again, stannic chloride is nonpolar but stannic fluoride, apparently on account of the smaller halogen atoms, is so polar as to form a high-melting solid.

[1] Cf. Rothmund, *Z. physik. Chem.*, **26**, 433 (1898).
[2] Cf. Debye, *Polar Molecules*, Chem. Catalog. Co., **1929.**
[3] Hildebrand, *Solubility*, Chem. Catalog. Co., **1924**, Chap. 8.
[4] Williams and Schwingel, *J. Am. Chem. Soc.*, **50**, 362 (1928).
[5] Dahms, *Ann. d. Physik. u. Chemie*, **54**, 486 (1895).
[6] Kremann, *Sitz. Akad. Wiss. Wien*, **117**, IIb, 569 (1908).
[7] Hildebrand, *Proc. Nat. Acad. Sci.*, **13**, 267 (1927); *J. Am. Chem. Soc.*, **51**, 66 (1929).
[8] Andrews, Lynn and Johnston, *J. Am. Chem. Soc.*, **48**, 1274 (1926).
[9] Desvergnes, *Mon. Sci.*, **15**, 149 (1925).
[10] Cf. *Tables ann. int. des const. et donnés numeriques*, **5**, 136 (1922).

---

*PENTAVALENT NITROGEN IN ORGANIC COMPOUNDS**

BY WILDER D. BANCROFT AND C. E. BARNETT

BAKER LABORATORY OF CHEMISTRY, CORNELL UNIVERSITY

Communicated February 24, 1930.

In connection with some work that we are doing on the constitution of the proteins, it became essential to know in what cases nitrogen in organic compounds will or will not add on hydrogen chloride gas stoichiometrically at room temperature and atmospheric pressure to give what is ordinarily called pentavalent nitrogen.

The following generalizations seem to cover most cases:

I. The tendency for a nitrogenous compound to react stoichiometrically with hydrogen chloride is increased when hydrogen is replaced by an alkyl group and decreased when hydrogen is replaced by a phenyl group.

II. Introduction of so-called negative radicals such as O, Cl, Br, NO_2, etc., decreases the tendency of the nitrogenous compounds to react stoichiometrically with hydrogen chloride. These radicals have most effect when attached directly to the nitrogen.

III. Carbonyl groups attached directly to the nitrogen decrease the tendency of the nitrogen to react stoichiometrically with hydrogen chloride. An aliphatic ethylene carbon linkage acts similarly. For convenience of discussion we shall call two carbonyl groups attached to nitrogen a diketo linkage, a CO group and a C:C aliphatic group attached to nitrogen a keto-ethylene linkage and two C:C aliphatic groups attached to nitrogen a diethylene linkage. Two keto linkages or one keto and one ethylene linkage will destroy the tendency to add hydrogen chloride unless compensated by the presence of alkyl groups. Since there are not many data at best, we usually mean methyl or ethyl groups when we speak of alkyl groups.

IV. Two or more nitrogens attached to the same carbon destroy the tendency to combine stoichiometrically with hydrogen chloride for all except one of the nitrogens, except insofar as this may be compensated by the introduction of alkyl groups.

V. In general, linkage of nitrogen to nitrogen seems to destroy the power of all but one of the nitrogens to combine stoichiometrically with hydrogen chloride, except insofar as this is compensated by the introduction of alkyl groups.

VI. Hydrogen chloride does not add direct to a nitrile nitrogen, an isonitrile nitrogen, or a cyanate nitrogen.

VII. If the nitrogen is already pentavalent through formation of an intramolecular salt, hydrogen chloride can only be taken up stoichiometrically in case it displaces the other acid.

VIII. Hydrogen chloride will add first to the nitrogen with which it will give the lowest dissociation pressure.

There is nothing new about these generalizations. They have either been made in some form by organic chemists or would be accepted by organic chemists as platitudes. This is the first time, however, that they have been brought together and tested.

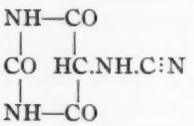
Under I we have the mono-, di- and tri-alkylamines combining stoichiometrically with hydrogen chloride. The ethylene amines add on as many hydrogen chlorides as there are nitrogens. Tetra-amidobenzene adds four hydrogen chlorides. Triphenylamine does not add hydrogen chloride, but tribenzylamine does. Pyridine, pyrrol, quinoline, naphthylamine and anthramine add hydrogen chloride but are weak bases. The corresponding hydrogenated compounds are stronger bases. Dimethylaniline is a stronger base than aniline.

Under II tribromaniline adds hydrogen chloride; but is a weak base. Nitro groups are more effective than the halogens in decreasing basicity, and dinitroaniline does not add on hydrogen chloride at all. Amidophenol is a weak base and so is phenylhydroxylamine.

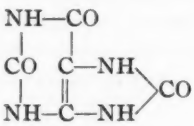
Under III we have acetamide and acetanilide reacting with hydrogen

chloride, while diacetamide and diacetanilide, succinimide and phthalimide do not react with hydrogen chloride because of the carbonyl groups. Parabanic acid, barbituric acid, dialuric acid, violuric acid, uramil and alloxan do not add hydrogen chloride stoichiometrically, though each has two or more nitrogen atoms in the molecule.

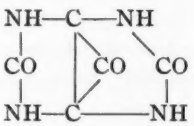
Three formulas have been suggested for uric acid:



Baeyer



Medicus



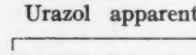
Fittig

The middle one, suggested by Medicus, has been universally adopted on chemical grounds. It is interesting to note that the same decision could have been made on the basis of generalization No. III and the fact that uric acid does not react stoichiometrically with hydrogen chloride. In the formula by Medicus the upper left-hand nitrogen cannot react with hydrogen chloride because of the diketo linkage and the other three because of the keto-ethylene linkage. In the Baeyer formula the diketo linkage bars out the left-hand nitrogens; but the first of the right-hand nitrogens would add hydrogen chloride stoichiometrically. In the Fittig formula two of the nitrogens would be barred under generalization No. IV; but uric acid should still be a diacid base.

Under IV we have one of the two nitrogens in urea adding hydrogen chloride stoichiometrically and one of the three nitrogens in guanidine. Guanazol, with six nitrogens, adds only two hydrogen chlorides. Hexamethylene tetramine is an apparent exception, because it has four nitrogens and adds four hydrogen chlorides stoichiometrically, although each carbon has three nitrogens attached to it. The apparently abnormal behavior is due to the fact that each nitrogen has three methylene groups attached to it, which compensate.

Under V we have phenyl hydrazine adding one hydrogen chloride, while ethyl hydrazine adds two, just as hydrazine itself does, the alkyl group making hydrazine a stronger base and the aryl group making it a weaker base.

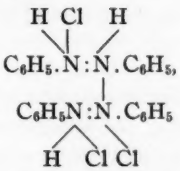
Urazol apparently forms no hydrochloride. It has the formula



and should add one hydrogen chloride unless we consider the —NH.NH. group functioning as a single nitrogen, in which case all of the nitrogens are barred because of the diketo linkage. Amidoguanidine has four nitrogens but only adds one hydrogen chloride because the fourth nitrogen is linked to one of the others. Aminotetrazol,

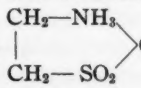
$H_2N.\overbrace{C:N.N:N.NH}$, has five nitrogens; but only one adds hydrogen chloride.

Azobenzene, $C_6H_5.N:N.C_6H_5$, is an apparent exception. It should add only one hydrogen chloride and it actually adds one and a half. The formula, however, is undoubtedly



which does not violate the rule, because the third hydrogen chloride does not add direct to any nitrogen and we are considering only this latter case.

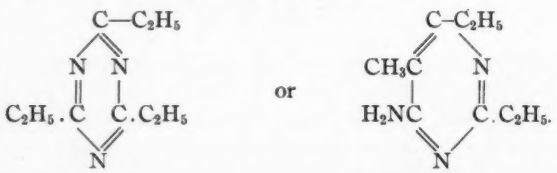
Under VI we find that hydrogen chloride does add reversibly and stoichiometrically to hydrocyanic acid, but not to give pentavalent nitrogen. It forms amino-formyl chloride, $H_2C(Cl):NH$. This substance would undoubtedly add hydrogen chloride to the nitrogen if it did not polymerize. The polymerization products of the nitriles, isonitriles, cyanates and isocyanates have not been studied with any satisfactory degree of care.

Under VII we find taurine, $NH_2CH_2.CH_2.SO_3H$, and hippuric acid, $NH_2C_6H_5.CO.NH.CH_2CO_2H$, which should each add one hydrogen chloride but neither does. In the case of taurine, everybody admits the existence of an intramolecular salt  . If the same explanation held for hippuric acid, the intramolecular salt would be decomposed if we formed the ethyl ester, and this compound should take up hydrogen chloride stoichiometrically, as indeed it does. Dr. P. A. Levene of the Rockefeller Institute was good enough to give us a sample of glycyl-glycyl glycine, for which we tender him our thanks. We expected that this compound would take up three hydrogen chlorides; but it only took two. On looking the matter up, we found that Emil Fischer had already expressed the opinion that this compound formed an intramolecular compound. This would account for the discrepancy.

Generalization VIII was put in almost as an after-thought to account for the behavior of succinimidine, which we have not yet tested ourselves. According to the literature this compound adds only one hydrogen chloride and the formula is written $HN:C.CH_2.CH_2.C:\overbrace{NH}$. If we write the formula this way, no more hydrogen chlorides can be added because the

other two nitrogens are barred under Generalization No. IV. If the first hydrogen chloride had added to one of the end nitrogens as written instead of to the connecting nitrogen, a second hydrogen chloride could have added to the nitrogen at the other end. If the literature is right—which we cannot guarantee—the connecting nitrogen must be the most basic of the three. A possible explanation is to be found under Generalization No. I, in that this nitrogen is the only one attached to two alkyl groups.

We have come across one other case which may belong here. According to Roscoe and Schorlemmer cyanethine or triethyl cyanuride is a mono-acid base. We can write the formula as



On the first formula, there might be only one hydrogen chloride taken up because the other two nitrogens are attached to carbons which are attached to the first nitrogen. On the pyrimidine formula there should be two hydrogen chlorides unless the first hydrogen chloride adds to the middle nitrogen, for which there is no obvious reason. If the experiments are right and if we must adopt the pyrimidine formula, as Sidgwick says, then Generalization No. VIII seems the only explanation.

The bearing of this work on the constitution of the proteins lies in the fact that only about thirty per cent in round numbers of the total nitrogen in gliadin combines stoichiometrically with hydrogen chloride, only about twenty per cent of the nitrogen in edestin and none of the nitrogen in zein.

If a protein were composed entirely of guanidine derivatives, that would leave only one-third of the total nitrogen available for hydrogen chloride; but that is not good enough for these three proteins. The so-called peptide linkage, —CONH—, adds on one hydrogen chloride for every nitrogen in all the simple cases tested. Dr. P. A. Levene furnished us with samples of glycyl glycine, which added two hydrogen chlorides, and of N-methyl-alanylglycyl glycine, which added three hydrogen chlorides. Professor John R. Johnson of the Department of Chemistry gave us a sample of glycyl-phenylalanine, which added two hydrogen chlorides, and reference has been made to Dr. Levene's glycyl-glycyl glycine, which added two hydrogen chlorides instead of three, presumably because of formation of an intramolecular salt.

The evidence so far indicates that zein contains practically no peptide

linkages and our problem is now to determine what nitrogen linkages it does contain.

The general results of this paper are as follows:

1. Eight generalizations have been given which cover pretty well the conditions under which organic nitrogen does or does not add hydrogen chloride stoichiometrically.

2. About seventy per cent of the total nitrogen in gliadin, eight per cent in edestin and one hundred per cent in zein does not add hydrogen chloride stoichiometrically.

3. All nitrogens having the peptide linkage, —CONH—, should add hydrogen chloride stoichiometrically.

4. There are apparently no peptide linkages in zein.

* This work is part of the programme now being carried out at Cornell University under a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

---

*THE TAUTOMERIC FORM OF MALIC ACID**

BY WILDER D. BANCROFT AND HERBERT L. DAVIS

BAKER LABORATORY OF CHEMISTRY, CORNELL UNIVERSITY

Communicated February 27, 1930

The specific rotations of the plane of polarized light of solutions of the so-called laevo-rotary malic acid and its sodium salt change with increasing dilation from being dextro-rotary to being laevo-rotary, as may be seen from the data of Woringer[1] and of Thomsen[2] in table 1.

TABLE 1

MALIC ACID			SODIUM MALATE		
C	M	$(a)_D$	C	M	$(a)_D$
63.2	4.71	+1.788°	87.76	4.93	+3.82°
45.68	3.41	+0.600	83.94	4.71	+3.35
36.00	2.68	+0.003	76.08	4.27	+2.37
26.73	1.99	−0.589	65.34	3.68	+0.68
20.92	1.56	−1.050	57.14	3.21	−0.78
15.87	1.18	−1.228	46.78	2.62	−2.54
10.65	0.79	−2.346	36.51	2.05	−4.24
7.52	0.56	−2.479	16.56	0.93	−6.94

In this table C is number of grams of solute per 100 cc. of solution, M is the number of mols of solute in 1000 cc. of solution and $(a)_D$ is the specific rotation calculated from the equation

$$(a)_D = \frac{a_D \times V}{l \times C}$$

where a_D is the observed rotation of the plane of polarized sodium light, l is the length of the tube in decimeters, and C is the number of grams of solute contained in the volume V of the solution.

Addition of salts, acids and bases tend to make the laevo-rotary acid more dextro-rotary. With rising temperature, the pure acid and also the solutions become more laevo-rotary. These changes cannot be due to electrolytic dissociation, because the effect of hydrochloric acid is quite marked up to relatively high concentrations and it would take relatively little acid to force back the dissociation of malic acid to a negligible value. Another reason is that we get a similar change with the concentration with malic ester in alcoholic solution.

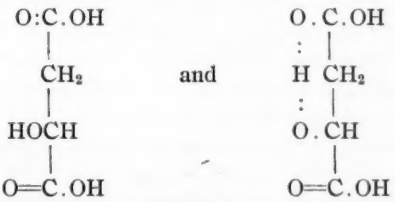
The changes cannot be due to changing polymerization, because freezing-point determinations show no signs of polymerization even at relatively high concentrations. More important are the observations by Lowry[3] on the tartrates that there is no direct relation between the degree of polymerization of a solute and its optical rotatory power. Lowry found that there are solutions of ethyl tartrate in various solvents in which the ester is polymerized increasingly with concentration and the specific rotation changes but little with change in concentration; while in other solvents the rotations change markedly with dilution although there is little or no evidence of polymerization.

The effects cannot be due to hydrates because we get them in alcohol and in acetone as well as in water and with the ester just as well as with the free acid. Rising temperature has the same qualitative effect as decreasing the concentration. Since rising temperature must decompose hydrates, we should have to conclude that there are fewer hydrates in dilute solutions than in more concentrated ones, which does not seem probable. Further, we can get the temperature effect with the anhydrous acid, where there can be no hydrates.

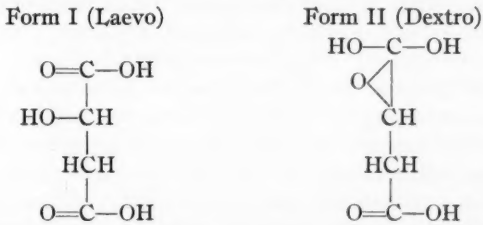
The change with the concentration cannot be due to the reversible formation and decomposition of a lactone of the ordinary type because we get the effect with ethyl malate as well as with malic acid. The change cannot be due to a reversible conversion of laevo-malic acid into dextro-malic acid, because then a solution of equivalent amounts of dextro- and laevo-malic acids would become optically active on addition of salts, acids and bases. Hydrochloric acid or sodium hydroxide imparts no activity to a solution of *dl*-malic acid. The changes on adding electrolytes to a solution of dextro-malic acid are equal and opposite in sign to the changes in laevo-malic acid under the same conditions.

We have thus eliminated ionization, polymerization, hydrate formation, β-lactone formation and formation of *d*-malic acid from *l*-malic acid as playing important parts in the change of optical rotation of malic acid with concentration, though some or most of these may play a minor part.

This leaves only a tautomeric change as a possibility and this is the view taken by Arndtsen, Armstrong and Walker, Lowry, de Malleman, Longchambon, Darmois and others. There is no question but that the changing rotation is due to the hydrogen of the hydroxyl group in the malic acid. If we replace it by a methyl group, the optical rotation is practically independent of the concentration. If we form a uranyl complex, we also eliminate the change of the rotation with the concentration pretty completely. Lowry has reached the same conclusion. Lowry would make the two formulas:



In the second formula the hydroxylic hydrogen is coördinated internally with the carbonyl group, the hydrogen being bivalent. We rather prefer to retain normal valences, and we, therefore, postulate that, in the case of laevo-malic acid, we have the following two formulas



Form I is the dilute solution form while Form II will predominate in concentrated solutions. In order to account for anomalous dispersion, we postulate that the two forms have different rotatory dispersion. There is, of course, no way of distinguishing between Form II and the Lowry type, because the two are merely different ways of saying the same thing.

Our Form II has two uncommon features. In the first place it contains two hydroxyl groups attached to the same carbon; but we have that in chloral hydrate. In the second place there is an ethylene oxide oxygen linkage. This might be called an alpha lactone with the water not split off. This formation of a ring structure is believed to account for the reversal of the sign of rotation. It is well known that the formation of the lactide from lactic acid, while not a lactone formation in the same

sense, does involve a change from a chain to a ring structure and does show the lactide to possess an opposite and remarkably enhanced rotation when compared with the original lactic acid. For many years Holmberg has been developing evidence to show that the hydrolysis of *l*-chlorosuccinic acid to *l*-malic acid involves an intermediate dextro-rotatory lactone, the three substances being of the same relative configuration. That such a dextro-rotatory lactone of malic acid should yield a laevo-rotatory malic acid on addition of water makes it seem reasonable that the reverse change might take place, causing the *l*-malic acid to become rotatory when a ring structure was produced.

The general results of this paper are as follows:

1. The changes in optical rotation and the anomalous dispersion of solutions of *l*-malic acid in solution are due to the presence of two tautomeric forms in dynamic equilibrium.

2. The laevo-rotatory form is the ordinary malic acid while the dextro-rotatory form is an ethylene oxide resulting from the migration of the hydrogen of the hydroxyl on the asymmetric carbon to the carbonyl oxygen of the carboxyl group.

3. It is immaterial whether one postulates the ethylene oxide formula or the coördinated form preferred by Lowry. The ethylene oxide formula has the merit of only requiring normal valences.

4. If the hydrogen of the hydroxyl on the asymmetric carbon is replaced by a methyl group which would be not expected to migrate readily, the optical rotation with changing concentration becomes negligible.

5. Since the phenomena occur just as well with the ester as with the free acid, the migrating hydrogen cannot be the carboxyl hydrogen.

6. The uranyl complex with malic acid does not show changing rotation with changing rotation.

7. The abnormal rotation and the anomalous dispersion cannot be accounted for satisfactorily on the basis of polymerization, electrolytic dissociation hydrates, lactones of the ordinary type, or of a structural reversal of the asymmetric carbon atom.

* This work is part of the programme now being carried out at Cornell University under a grant from the Heckscher Foundation for the Advancement of Research established by August Heckscher at Cornell University.

[1] Woringer, *Z. physik. Chem.*, **36**, 636 (1901).

[2] Thomsen, *J. prakt. Chem.*, **143**, 753 (1887).

[3] Lowry and Cutter, *J. Chem. Soc.*, **125**, 1465 (1924).

ON THE OVERCONVERGENCE OF SEQUENCES OF POLYNOMIALS OF BEST APPROXIMATION

BY J. L. WALSH

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated February 26, 1930

It is the object of this note to state briefly some results, in part new, the proof of which will shortly appear elsewhere. Detailed references to the literature will be given at that time. The interest of the present results lies first in the unification of the method of proof, for they all follow from a single theorem (see Walsh, *Münchner Berichte*, 1926, pp. 223–229) on degree of approximation; and second in their generality, for, broadly speaking, the previous results have not included the present measures of approximation nor the general regions treated here.

Let C be an arbitrary closed limited region in the $z(= x + iy)$-plane and let D denote the region consisting of all points which can be joined to the point at infinity by Jordan arcs which do not meet C. Let C_R denote the image in the z-plane of the circle $|w| = R > 1$ when D is so mapped onto the exterior of the circle $|w| = 1$ in the w-plane that the points at infinity correspond to each other. Then under certain restrictions to be mentioned, *if the function* $f(z)$ [*or* $u(x, y)$] *is analytic* [*or harmonic*] *interior to* C_R, *then the sequence of polynomials in* z [*or harmonic polynomials*] *of best approximation to* $f(z)$ [*or* $u(x, y)$] *on* C *converges to* $f(z)$ [*or* $u(x, y)$] *interior to* C_R, *uniformly on any closed point set interior to* C_R. Here *best approximation* may be (1) in the sense of Tchebycheff, C being an arbitrary limited region; (2) in the sense of least p-th powers ($p \geqq 1$) measured by integration over the boundary of C; this boundary is assumed a rectifiable Jordan curve for approximation to $f(z)$, a Jordan curve with continuous curvature for approximation to $u(x, y)$; (3) in the sense of least p-th powers on the circumference γ: $|w| = 1$, where C is an arbitrary limited simply connected region and is mapped conformally onto the interior of γ; (4) in the sense of least p-th powers measured over the area of C, the region C being an arbitrary limited region.

More restricted results are proved if $f(z)$ [or $u(x, y)$] is not assumed analytic [or harmonic] in C_R, but is assumed analytic [or harmonic] interior to C, continuous in the closed region. In each case the sequence of polynomials of best approximation to $f(z)$ [or $u(x, y)$] on C converges to $f(z)$ [or $u(x, y)$] interior to C, uniformly on an arbitrary closed point set interior to C, provided merely that $f(z)$ [or $u(x, y)$] can be uniformly approximated as closely as desired in the closed region C by some polynomial in z [or harmonic polynomial].

SUMMARY OF RESULTS AND PROOFS ON FERMAT'S LAST THEOREM

(*Fifth Paper*)

BY H. S. VANDIVER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS

Communicated March 10, 1930

In a recent paper[1] the writer proved a number of results concerning Fermat's Last Theorem. In the present article I shall give a number of sidelights and comments on the contents of that paper as well as some extensions of the theorems therein.

One method the writer used in order to check the accuracy of the results and proofs in the paper just mentioned was to attempt to find extensions of these results without introducing essentially new ideas. In trying this there were often found in lieu of extensions, new proofs of the old results or variations of the original proofs and a number of these will be explained here. The results in T mainly concerned the second case of Fermat's Last Theorem, that is, when

$$x^l + y^l + z^l = 0 \tag{1}$$

has solutions in rational integers prime to each other, but $xyz \equiv 0 \pmod{l}$. The writer recently attempted to apply some of these methods to the first case of the theorem, that is, when $xyz \not\equiv 0 \pmod{l}$ in (1). A number of new results were found including the following Theorem. *If (1) is satisfied in case I then* $B_{nl} \equiv 0$ *(Mod* l^3*), for* $n = (l-3)/2$, $(l-5)/2$, $(l-7)/2$ *and* $(l-9)/2$ *where, following the notation of T, the B's are the Bernoulli numbers.*

Comparing this result with another in a former paper,[2] and using the relation[3], if $l_1 = (l-1)/2$,

$$B'_{n+xl_1} \equiv xB'_{n+l_1} - (x-1)B'_n \pmod{l^2}, \tag{2}$$

$$B'_n = \frac{(-1)^n B_n}{n},$$

we obtain

THEOREM VIII. *If the relation* (1) *is satisfied in Case I, then*

$$B_{n_1 l_1 - 1} \equiv B_{n_2 l_2 - 2} \equiv B_{n_3 l_1 - 3} \equiv B_{n_4 l_1 - 4} \equiv 0 \pmod{l^2},$$

where n_1, n_2, n_3 *and* n_4 *each range independently over all positive integral values,*[4] *the B's being the Bernoulli numbers:* $B_1 = 1/6$, $B_2 = 1/30$, etc.

These methods and results enable us, in most cases, to obtain criteria for the solution (1) by extending the criteria found for Case I of (1) to Case II. For example, in view of these new results on Case I, we enunciate

Theorem I, of T (which is the same as Theorem I of the fourth paper under the present title) without the restriction to Case II. Theorem V of T is an example of how we may obtain criteria for the solution of (1) by using criteria for Case I and then adding an extra condition which provides for Case II. In this way a number of new results concerning (1) without restriction (except that x, y and z are rational integers none zero) may be obtained from the methods and results in some former papers[4] of the writers. Further details along this line I hope to give in another paper.

Another example of a result already proved for the first case which also holds for the second case is the following:

THEOREM IX. If the number of independent generators of the group containing all the C classes in the irregular class group of the field $k(\zeta)$; $e^{2i\pi/l} = \zeta$, is equal to the number of independent generators of the group of c classes (these terms are all defined in T, p. 625), then (1) is impossible in rational integers none zero.

For Case I this was proved by Hecke.[5] It may be proved for Case II by observing, as did Hecke, that in a field of the type mentioned there exists no singular primary unit. We may start the same as in the proof of Theorem II of T and obtain the relation (12) given there, but since there exists no singular primary unit in our field, then n_a^2/n_b^2 in (12) is an l-th power. We may then apply immediately the method of infinite descent described in the proof in T of Theorem I.

In the proof of Theorem I of T the second assumption in the statement of the theorem might have been replaced, as is known[6] by the assumption that if there exists a unit in $k(\zeta)$ which is of the form $a^l(\text{mod } l^2)$, with a a rational integer than it is the l-th power a unit in $k(\zeta)$. In this connection I have proved the

THEOREM A. *If the second factor of the class number of $K(\zeta)$ is prime to l then a necessary and sufficient condition that a unit of the form $a^l(\text{mod } l^2)$, where a is a rational integer, is a singular unit, is that $B_{nl} \equiv 0(\text{mod } l^3)$ for some n in the set $1, 2, \ldots, l_1$.*

This condition was proved necessary in T, Lemma II. To prove it sufficient, we first employ a theorem which I shall prove in another paper,[7] to the effect that a necessary and sufficient condition that the second factor of the class number of $k(\zeta)$ is prime to l is that at least one of the units

$$E_n = \prod_{i=0}^{l_1-1} \epsilon(\zeta^{r^i})^{r^{-2in}},$$

$$n = 1, 2, \ldots, l_1;$$

$$\epsilon = \left(\frac{(1-\zeta^r)(1-\zeta^{-r})}{(1-\zeta)(1-\zeta^{-1})}\right)^{\frac{1}{2}};$$

r is a primitive root of l and $\epsilon(\zeta^{r^i})$ is the unit obtained from $\epsilon(\zeta)$ by the substitution (ζ/ζ^{r^i}) is the l-th power of a unit in $k(\zeta)$. In view of this and another theorem due to the writer,[8] an equation of the type

$$E_1^{a_1} E_2^{a_2} \ldots E_{l_1}^{a_{l_1}} = \gamma^l$$

cannot exist unless all the a's are divisible by l. We may then show that if $B_{nl} = 0 \pmod{l^3}$, then the exponents c in

$$E_n E_1^{c_1 l} E_2^{c_2 l} \ldots E_{l_1}^{c_{l_1} l} = \delta$$

may be selected so that δ is of the form $a^l \pmod{l^2}$, and by what precedes we know that δ is not an l-th power. To effect this we write, if $\lambda = 1 - \zeta$

$$\delta = a^l + \theta_i \lambda^i; \quad l < i < 2l \tag{3}$$

and show that $\theta_i \equiv 0 \pmod{\lambda}$ successively for each value of i in the limits mentioned. From (2) we obtain an identity in e^v as in the proof of Lemma II in T, and we consider

$$\left[\frac{d^s \log \delta(e^v)}{dv^s}\right]_{v=0}$$

for various values of s; $l < s < 2l$ and also use for comparison some of these differential coefficients for $s < l$; in particular if $i = 2n$ we need to compare them for $2n$ and $l - 1 + 2n$ and use the relation (2) between Bernoulli numbers, noting that $B_{nl} \equiv 0 \pmod{l^3}$.

It may be observed that in the proofs of all the Theorems I to V of T that the method of infinite descent described in Kummer's *Memoir of 1857* was employed. However, new proofs of Theorems VI and V constructed by the writer, in which, after the formulas (26), (26a), (9) and (9a) are obtained the procedure is quite different from anything used by Kummer in his paper just cited. The method of descent employed by the writer depends on reducing the exponent of λ in the successive relations of the type (9a) instead of reducing the number of distinct ideal factors in certain integers appearing in the successive relations obtained, as did Kummer.

A new proof of the fundamental Lemma I of T may be derived by using the theorem already referred to giving a necessary and sufficient condition, that the second factor of the class number of $k(\zeta)$ is divisible by l and employing Corollary I of a previous paper.[9] This proof is quite different from that given in T and the other proofs referred to there, as it does not involve the use of laws of reciprocity or the theory of the class-field.

The proof of Theorem I of T evidently holds also for x, y and z integers

in the field defined by $\zeta + \zeta^{-1}$ and prime to each other. Theorems II and IV also hold under like condition. Possibly by an extension of the method by Hilbert[10] for the proof that, if l is a regular prime,

$$\alpha^l + \beta^l + \gamma^l = 0 \tag{4}$$

is impossible for any non-zero integers in the field $k(\zeta)$ we might extend Theorems I, II and IV of T to x, y and z any non-zero integers in $k(\zeta + \zeta^{-1})$, but I have not attempted to carry out the details.[11]

The proof of Theorem IV of T is quite complicated and I have many times attempted to find variations of it. The most difficult part of the proof which involves the derivation of (27) to (30) inclusive and up to the end of the proof that $\omega + \theta \equiv 0 \pmod{p}$ may be replaced by another argument if we observe that, using $\omega + \theta \equiv 0 \pmod{l^2}$,

$$\left[\frac{d^{2s+1} \log \omega(e^v)}{dv^{2s+1}}\right]_{v=0} \equiv 0 \pmod{l^2}$$

we have by changing (26) of T into an identity in e^v

$$\left[\frac{d^{2s+1} \log \rho_a(e^v)}{dv^{2s+1}}\right]_{v=0} \equiv 0 \pmod{l}.$$

From the results in a forthcoming paper[12] we find that ($\mathfrak{P}$ used as in *Th IVa*)

$$\left\{\frac{E_n}{\mathfrak{P}}\right\} = 1;\ n = 1, 2, \ldots l_1$$

which proves that $\omega + \theta \equiv 0 \pmod{p}$. It is possible to give two other variations of the proof in T of this.

The Theorem I of T involves two assumptions as follows:

1. The second factor of the class number of k (ζ) is prime to l.
2. None of the Bernoulli numbers B_{nl} is divisible by l^3; $n = 1, 2, \ldots, l_1$.

In Theorems IV and V we do not need the second assumption just mentioned, but the first assumption is included in the assumption stated in both IV and V since if

$$\left\{\frac{E_n}{\mathfrak{P}}\right\} \neq 1$$

then E_n is not the l-th power of a unit in $k(\zeta)$. However, it is possible to prove two theorems which include Form V and which do not involve, apparently, either of the assumptions of Theorem I. They are:

THEOREM IV*a*. *Under the following assumption:*

If $\eta_{a_1} \eta_{a_2} \ldots \eta_{a_s}$ *is a system of units in* $k(\zeta)$ *such that every singular*

primary unit may be written in the form $\eta_{a_1}^{d_1}, \eta_{a_2}^{d_2}, \ldots \eta_{a_s}^{d_s}\delta^l$, *with* δ *a unit in* $k(\zeta)$, *the d's integers and*

$$\left\{\frac{\eta_{a_i}}{\mathfrak{P}}\right\} \neq 1;\ i = 1, 2, \ldots, s;$$

$\mathfrak{P}$ *is a prime ideal divisor of* p; p *is a prime* $< l^2 - l$ *of the form* 1 (*mod* l); *the relation* (2) *is impossible in Case II.*

THEOREM V*a*. *Under the following assumptions;*

1. *There exists a rational prime integer* p *such that the congruence*

$$u^l + v^l + w^l \equiv 0 \pmod{p}$$

has no solution; u, v *and* w *all rational integers prime to* p *and* $p \not\equiv 1 \pmod{l^2}$.

2. *The relation*

$$\left\{\frac{\eta_{a_i}}{\mathfrak{P}}\right\} \neq 1$$

holds where i *ranges over the values* 1, 2, . . ., s; $\eta_{a_1}, \eta_{a_2}, \ldots, \eta_{a_s}$ *form a basis as defined in Theorem IV*a *and* $\mathfrak{P}$ *is a prime ideal divisor of* p; *the relation* (2) *is impossible in rational integers none zero.*

The proofs of these theorems are obtained by extensions of the methods applied in the proofs of Theorems IV and V in *T*.

Note that in *T*, the computation mentioned on the last two pages in checking Theorems I and IV, involve finding

$$\left\{\frac{E_n}{\mathfrak{P}}\right\}$$

for special values of n and p. It was found in each case that the *smallest* prime of the form $1 + tl$ gave a value such that

$$\left\{\frac{E_n}{\mathfrak{P}}\right\} \neq 1.$$

Since E_n is a singular primary number in $k(\zeta)$ then the above relation can hold only if $\mathfrak{P}$ belongs to the irregular class group of $k(\zeta)$, then, in general, the question depends on the distribution of ideals of the first degree in the various classes forming the irregular class group. By the use of a theorem due to Pollaczek[13] it is possible to show that there exist some fields $k(\zeta)$ which possess classes in the irregular class group in which all the ideals therein are of the first degree. If an ideal is not in the first degree, then there exists a substitution $s = (\zeta/\zeta^r)$ such that $\mathfrak{P}s^n = \mathfrak{P}$; n less than and a divisor of $l - 1$; hence by the result of Pollaczek we have that

$$\mathfrak{P}^s \sim \mathfrak{P}^{r^{l-2a_i}};\ \mathfrak{P}^{s^n} \sim \mathfrak{P}^{r^{n(l-2a_i)}} \sim \mathfrak{P}$$

if $\mathfrak{P}$ belongs to a basis class defined by

$$C^{s^{l-2ai}} = 1,$$

and

$$r^{n(l-2a_i)} \equiv 1 \pmod{l}.$$

This is only possible if $l - 2a_i$ has a factor in common with $l - 1$. In a properly irregular cyclotomic field[14] the integer a_i must be such that

$$B_{a_i} \equiv 0 \pmod{l}.$$

Hence we may state the

THEOREM B. *For there to exist an ideal not of the first degree in the basis classes C of the irregular class group of $k(\zeta)$ defined by*

$$C^{s-r^{l-2a}} = 1,$$

it is necessary that $l - 2a$ and $l - 1$ have a common factor.

Applying this result together with the statement just made concerning the B's to the fields defined by $l = 37, 59, 67, 101, 103, 131, 149$ and 157, we find that all ideals in the irregular class group in each of these fields are of the first degree except possibly for $l = 67$.

In T, Theorem VII stated that $x^n + y^n = z^n$ is impossible in rational integers none zero for $2 < n < 211$. This theorem has recently been extended by the writer and his assistants to $n < 269$. First, all primes l; $211 < 1 < 269$, were tested as to being regular. All were found regular between these limits except 233, 257 and 263. Where $l = 233$, it was found that $B_{42} \equiv 0 \pmod{l}$, but this is true of no other B's in the set, $B_1, B_2, B_3, \ldots, B_{l_1}$. For $l = 257$, it was found that $B_{82} \equiv 0 \pmod{l}$ but this relation did not hold for any other B in the above set. Similarly, where $l = 263$, $B_{50} \equiv 0 \pmod{l}$. The divisibility of these Bernoulli numbers was also checked by using the actual values of the numbers given by Adams.[15] These computations were carried out by Miss Elizabeth M. Badger and are included in her M.A. Thesis at the University of Texas. Theorem II was applied to the three non-regular primes mentioned, the computations for the verification of assumption II being carried out by Mr. J. A. Clack, graduate student at the University of Texas. He found, using the notation of T, p. 639, that $A_{42} \equiv 26.233 \pmod{233^2}$; $A_{18} \equiv 186.257 \pmod{257^2}$; $A_{50} \equiv 162.263 \pmod{263^2}$. Hence Fermat's last theorem is established for all exponents $2 < l < 269$. Obviously, Theorem III of T also yields proofs for $l = 233$ and 257. The other theorems have not been tested for these three cases. We, therefore, state:

THEOREM VI*a*. *The equation*

$$\alpha^l + \beta^l + \gamma^l = 0$$

is impossible for all odd prime exponents $l < 269$ *excepting possibly* $l = 37$, 59, 67, 101, 103, 131, 149, 157, 233, 257 *and* 263 *if* α, β *and* γ *are integers in the field* $k(\zeta)$ *none zero.*

THEOREM VII*a*. *The equation*

$$x^n + y^n = z^n$$

is impossible in rational integers none zero if

$$2 < n < 269.$$

As a pattern for his computations on the primes $l = 233$, etc., Mr. Clack considered the prime $l = 67$ and found that $B'_{29\cdot 67} \equiv 67.13 \pmod{67^2}$, checking the result found by Mrs. Williams; cf. *T.*, p. 641; B' as here used is also defined there. Computations are now under way with a view to extending Theorem VII*a* to $n < 300$.

[1] *Trans. Amer. Math. Soc.*, **31**, 1929 (613–42). Will be referred to as *T*.

[2] *Ann. Math.*, **26**, 1925 (231).

[3] Pollaczek, *Math. Zeits.*, **21**, 1924 (28).

[4] *Proc. Nat. Acad. Sci.*, **12**, 1926 (106–109); *Trans. Am. Math. Soc.*, **28**, 1926 (554–560); **29**, 154–162 (1927).

[5] Göttingen Nachr., *Math.-Physik. Kl.*, 1910 (420–424).

[6] *Bull. Nat. Res. Coun.*, Feb., 1928, **62**, Report on Algebraic Numbers—II, p. 45; *Proc. Nat. Acad. Sci.*, **12**, 1926 (767–772); *T*, p. 623, relation (10).

[7] *Proc. Nat. Acad. Sci.*, **16**, 1930.

[8] *Bull. Amer. Math. Soc.*, **35**, 1929 (323–335).

[9] *Proc. Nat. Acad. Sci.*, **15**, 1929 (207).

[10] *Jahresber. Deut. Math. Verein.*, 1897 (517–525).

[11] Professor E. Noether kindly communicated to me the information that Dedekind in a letter to Frobenius (reference not yet located in the literature) pointed out that Kummer's proof that (4) is impossible if α, β and γ are integers in the field $k(\zeta)$ and l a regular prime did not cover the case where α, β and γ had a common ideal factor in the field. Dedekind in the same letter gave a method for disposing of the difficulty which is similar to that employed by Hilbert (loc. cit.) for the same purpose. In the report on Algebraic Numbers, II (loc. cit.) the writer stated that Kummer proved (4) impossible only when α, β and γ are prime each to each.

[12] *Trans. Amer. Math. Soc.*, **31**, 1930, entitled, "On Power Characters of Singular Integers in a Properly Irregular Cyclotomic Field."

[13] *Math. Zeit.*, **21** (1924), p. 11 and p. 21.

[14] To appear in *Trans. Amer. Math. Soc.*, 1930, entitled, "On the Composition of the Ideal Class Group in an Irregular Cyclotomic Field."

[15] *Crelle*, **85**, 269 (1878).

GROUPS GENERATED BY TWO GIVEN GROUPS

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated March 15, 1930

If H_1 and H_2 are any two given groups of orders h_1 and h_2, respectively, and if K is their cross-cut, then the order g of the group G generated by H_1 and H_2 cannot be less than $h_1h_{2/k}$, where k is the order of K, and when $g > h_1h_{2/k}$ it must exceed this number by a number which is divisible by each of the two numbers h_1 and h_2. A proof of these statements results directly from the fact that if $1, s_2, \ldots, s_{h_1}$, represent the operators of H_1 and if $t_1, t_2, \ldots, t_l$ represent a set of operators of H_2 such that one and only one of them belongs to each co-set of H_2 with respect to K, then all the operators of the following rectangle are distinct:

$$\begin{array}{cccc} 1 & s_2 & \ldots & s_{h_1} \\ t_1 & s_2t_1 & \ldots & s_{h_1}t_1 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ t_l & s_2t_l & \ldots & s_{h_1}t_l \end{array}$$

in view of the fact that no t could appear in a row which precedes the one which involves this t. If G involves any additional operators their number must be divisible by both h_1 and h_2, since g is divisible by these two numbers and $h_1(l + 1)$ has the same property.

We shall first consider some necessary and sufficient conditions that g has the minimal value $h_1h_{2/k}$. It is well known that a sufficient condition is that at least one of the two subgroups H_1, H_2 is transformed into itself by the other and hence is invariant under G. It therefore results that the necessary and sufficient conditions that the order of G is $h_1h_{2/k}$ are direct generalizations of the important concept of invariant subgroup. Suppose that every operator of a set of generators of H_2 transforms every operator of H_1 into an operator of H_1 multiplied on the right by an operator of H_2. When this condition is satisfied it is easy to see that the operators of the given rectangle constitute a group, and hence this rectangle can also be written in a form in which the first row is composed of the operators of H_2 and the new t's are selected from the co-sets of H_1 with respect to K in the same way as the given t's were selected from the co-sets of H_2 with respect to K. Hence we have established the following theorem: *If a set of independent generators of the group H_2 transforms all the operators of a group H_1 into operators of H_1 multiplied on the right by operators of H_2 then must also all the operators of every set of independent generators of H_1 transform all the operators of*

H_2 into operators of H_2 multiplied on the right by operators of H_1, and the order of the group generated by H_1 and H_2 must be h_1h_2/k, where h_1 and h_2 are the orders of H_1 and H_2, respectively, and k is the order of their cross-cut.

Instead of saying "multiplied on the right" in this theorem we could obviously also say "multiplied on the left." We could also replace this condition by the following: The square of every operator of the given rectangle which has for its t constituent an operator of a fixed set of independent generators of H_2 is contained in this rectangle. In fact, this condition implies that a set of independent generators of H_2, transforms all the operators of H_1 into operators of H_1 multiplied on the right by operators of H_2, and conversely the latter condition implies the former. It should be emphasized that these conditions are reciprocal. That is, if it is possible to find one set of independent generators of H_2 such that the squares of all the products obtained by multiplying on the right the operators of H_1 successively by these independent generators are equal to the operators of H_1 multiplied on the right by operators of H_2, then the same is true not only for every set of independent generators of H_2 but also for every set of independent generators of H_1 as regards the group H_2. When H_1 and H_2 satisfy these conditions they are said to be *permutable*.[1]

While the concept of permutable subgroups was introduced early into group theory it has not as yet found extensive applications. In this respect it differs widely from the concept of invariant subgroup notwithstanding the fact that it is more general than the latter. The following theorem may serve to illustrate its generality: *If a non-regular transitive substitution group of degree n contains a transitive subgroup of this degree then this subgroup and any one of the subgroups composed of all the substitutions which omit one letter of the former group must be a pair of generating permutable subgroups of this group.* A proof of this theorem results directly from the facts that the two subgroups in question generate the entire group and that their cross-cut is composed of *all* the substitutions which omit one letter of the latter subgroup and are contained therein. It results directly from this theorem that every symmetric group whose order exceeds 2 and every alternating group whose order exceeds 3 must be generated by two of its permutable subgroups. Groups which involve at least one pair of generating permutable proper subgroups have been called decomposable.[2] It should be noted that a necessary and sufficient condition that the subgroup composed of all the substitution which omit one letter of a transitive group of degree n belongs to a pair of permutable subgroups is that the second subgroup is also a transitive group of degree n.

Every cyclic group whose order is a power of a prime number is obviously indecomposable. Suppose that G is any other group which has a solvable

quotient group. It is then known that G involves an invariant subgroup of prime index p. This subgroup and any Sylow subgroup of G whose order is a power of p must constitute a pair of generating permutable subgroups of G since their cross-cut must be of index p under the latter group. That is, *the cyclic prime power groups are composed of all the possible indecomposable groups which have solvable quotient groups*. A necessary and sufficient condition that two subgroups of a given group constitute a pair of generating permutable subgroups is that the index of one of these subgroups under the group is equal to the index of their cross-cut under the other subgroup. If two subgroups are conjugate it is known that the latter index is smaller than the former and hence it results, in particular, that *two generating permutable subgroups of a group cannot be conjugate under this group.*

If H_1 and H_2 are two generating permutable subgroups of G any one of the following three conditions may be satisfied. It may be that neither of them is invariant under G, that one of them is invariant under G while the other does not have this property, that each of them is invariant under G. A necessary and sufficient condition that the second of these conditions is satisfied is that all of the commutators which involve an element from each of these subgroups are found in one of these subgroups, while a necessary and sufficient condition that both of the subgroups H_1 and H_2 are invariant under G is that all of these commutators appear in the cross-cut of H_1 and H_2. When this condition is satisfied this cross-cut is also an invariant subgroup of G and the corresponding quotient group is the direct product of its two subgroups corresponding to H_1 and H_2, respectively.

The present article is a generalization of section 25 in the *Theory and Applications of Finite Groups* by Miller, Blichfeldt, and Dickson, 1916, where the construction of groups with invariant subgroups is considered. When H_1 and H_2 satisfy the conditions noted in the theorem stated in the second paragraph of this article the order of the group generated by H_1 and H_2 is the same as when H_1 is an invariant subgroup of G. The present article has also close contact with an article by E. Maillet which appeared under the heading "Sur les groupes échangeables et les groupes décomposables," *Bulletin de la Société Mathématique de France*, volume 28 (1900), page 7. It seems that no one has as yet given an example of a group which is indecomposable but does not belong to the category of cyclic prime power groups noted above.

In connection with the theorem noted in the fourth paragraph of this article it may be of interest to add that a necessary and sufficient condition that a Sylow subgroup of a transitive group of degree n is transitive is that n divides the order of this subgroup. Hence it results that a necessary and sufficient condition that the subgroup composed of all the sub-

stitutions which omit one letter of a transitive group is one of a pair of permutable subgroups in which the second subgroup is a Sylow subgroup is that the degree of this transitive group divides the order of this Sylow subgroup. As a subgroup of an intransitive group of degree n cannot be transitive on these n letters it results from this theorem that the Sylow subgroups appear only as transitive subgroups of a substitution group of degree n when the latter group is transitive and when its degree divides the order of the Sylow subgroups. In particular, when the degree of a transitive group is a power of a prime number its subgroup composed of all its substitutions which omit one letter is always one of a pair of permutable subgroups.

A sufficient condition that two groups each of which is generated by a pair of permutable subgroups are simply isomorphic is that a simple isomorphism can be established between these pairs such that the second subgroup of each pair transforms the corresponding operators of the first subgroups into corresponding operators of these subgroups multiplied on the right by corresponding operators of the second subgroups. This is clearly a generalization of the well-known theorem for establishing a simple isomorphism between two groups by means of a simple isomorphism between their invariant subgroups.[3]

[1] A. L. Cauchy, *Exercises d'Analyse*, **3** (1844), p. 227.

[2] The term decomposable was used by E. Maillet, *Bull. Soc. Math. France* **24** (1896), p. 85. G. Frobenius called these groups zerlegbar, *Sitzgsb. Akad. Berlin*, 1895, p. 166.

[3] Miller, Blichfeldt, Dickson, *Finite Groups*, **1916**, p. 73.

ON THE STRUCTURE OF PURE RIEMANN MATRICES WITH NON-COMMUTATIVE MULTIPLICATION ALGEBRAS

By A. Adrian Albert

Department of Mathematics, Columbia University

Communicated March 7, 1930

The chief outstanding problem in the theory of Riemann matrices* is the determination of all pure Riemann matrices. It is the purpose of this paper to give certain basic theorems on the structure of such matrices, and, in particular, the case where the multiplication algebra is a known normal division algebra.

1. *Omega Matrices over F.*—Let F be any field of complex numbers and Ω be any m-rowed square matrix of complex elements in the form

$$\Omega = \left\| \begin{matrix} \omega_1 \\ \omega_2 \end{matrix} \right\|$$

where ω_1 has p rows and $\bar{\omega}_2$ has m-p rows. The matrix Ω is called an *Omega matrix over* F if there exists an m-rowed skew-Hermitian matrix C with elements in F such that

$$\omega_1 C \omega_2' = 0 \tag{1}$$

$$H_1 = \omega_1 i C \bar{\omega}_1' , \; H_2 = \omega_2 i C \bar{\omega}_2' , \; (i = \sqrt{-1}), \tag{2}$$

are positive, definite, non-singular Hermitian matrices. An m-rowed square matrix A with elements in F is said to define a multiplication a of Ω if there exists a p-rowed complex square matrix α_1 and an $(m$-$p)$-rowed complex square matrix α_2 such that

$$\alpha_1 \omega_1 = \omega_1 A , \; \bar{\alpha}_2 \bar{\omega}_2 = \bar{\omega}_2 A. \tag{3}$$

Matrices A as above are called *projectivities of* Ω. The set of all multiplications of Ω is a linear associative algebra D over F, the set of all projectivities of Ω forms an m-rowed *first representation* D_1 of D as an algebra of square matrices with elements in F, and the set of matrices α_1 forms a complex p-rowed *second representation* D_2 of D.

Two Omega matrices, Ω_1 and Ω_2, are called *isomorphic in* F if there exists a non-singular m-rowed square matrix G with elements in F and two non-singular complex matrices γ_1 and γ_2 such that

$$\Omega_2 = \left\| \begin{matrix} \gamma_1 & 0 \\ 0 & \bar{\gamma}_2 \end{matrix} \right\| \Omega_1 G.$$

The multiplication algebras of isomorphic Omega matrices are the same and hence we need to consider *classes of isomorphic matrices*. When F is a real field, $m = 2p$, $\omega_1 = \omega_2 = \omega$ and Ω composed of ω and its conjugate matrix is an Omega matrix, the matrix ω is called a Riemann matrix over F. In particular when F is the field of all rational numbers Riemann matrices over F become ordinary Riemann matrices. Omega matrices and Riemann matrices over a field occur as sub-matrices of ordinary Riemann matrices and it is necessary to define and consider them.

2. *Basic Theorems for Pure Riemann Matrices.*—THEOREM 1. *Let B be any division algebra over a field of complex numbers F and let B_1 and B_{11} be any two representations of B as an algebra of m-rowed square matrices with elements in F such that the modulus and zero element of B are represented in both B_1 and B_{11} by $I(m)$, the m-rowed identity matrix, and $0(m)$, the m-rowed square zero matrix, respectively. Then there exists a non-singular m-rowed square matrix P with elements in F such that if any x of B is represented by X_1 in B_1 and X_{11} in B_{11} then*

$$X_{11} = P X_1 P^{-1}.$$

The above theorem enables us to utilize any desired first representation

D_1 of D for a Riemann matrix ω by passing to ωP^{-1} isomorphic in F to ω. We also have

THEOREM 2. *The structure of all pure Riemann matrices is determined by the structure of all Riemann matrices over a real field F with multiplication algebras of the following three types:*

I. *A normal division algebra B of order n^2 over F containing a quantity a of grade n with respect to F and with minimum equation having all real roots $\alpha_1, \ldots, \alpha_n$.*

II. *The same as* I *but with $\alpha_1, \ldots, \alpha_n$ all imaginary, $n = 2n'$, a polynomial $\theta(\alpha_1)$ with coefficients in F existing and having the property that if $\alpha_{j+n} \equiv \alpha_j$ then $\bar{\alpha}_j = \bar{\alpha}_{j+n'} = \theta(\bar{\alpha}_j)$, $\theta^2(\alpha_j) = \alpha_j$ $(j = 1, \ldots, n)$.*

III. *A division algebra B of order $2n^2$ over F with B a normal division algebra of order n^2 over a quadratic field $F(q)$, $q^2 = \epsilon$ in F. Moreover, B contains a quantity a of grade n with respect to both F and $F(q)$ and such that the roots of its minimum equation with respect to these fields are all real.*

THEOREM 3. *Let B be a normal division algebra over any complex field K and let B_{21} and B_{22} be any two representations of B as an algebra of complex p-rowed square matrices with the modulus and zero element of B represented in both B_{21} and B_{22} by $I(p)$ and $0(p)$, respectively. Then there exists a non-singular complex p-rowed square matrix π such that if x of B is represented in B_{21} by ξ_1 and in B_{22} by ξ_2, then*

$$\xi_2 = \pi \xi_1 \pi^{-1}.$$

Hence when the multiplication algebra of ω is either I or II we may take any desired second representation of D. In the case III it is easily shown that n divides p, $p = np'$, $p' = r_1 + r_2$ where r_1 and r_2 are arbitrary positive or zero integers satisfying the condition that their sum is p', and q has the second representation

$$\rho = \left\| \begin{matrix} \sqrt{\epsilon}\, I\,(r_1) & 0 \\ 0 & -\sqrt{\epsilon}\, I\,(r_2) \end{matrix} \right\|. \tag{4}$$

We then have

THEOREM 4. *Let B be a normal division algebra over $F(q)$ where F is a real field and $q^2 = \epsilon < 0$ in F. Suppose that B_{21} and B_{22} are any two complex representations of B in both of which the modulus, the zero element, and q of B are represented respectively by $I(p)$, $0(p)$, and the same ρ of* (4). *Then there exists a complex p-rowed non-singular matrix π such that if x of B is represented by ξ, in B_{21} and by ξ_2 in B_{22}, then*

$$\xi_2 = \pi \xi_1 \pi^{-1}.$$

Hence, apart from the arbitrary r_1 and r_2 we may take any desired second representation of D.

3. *The Case of Known Division Algebras.*—Let B be an associative

normal division algebra in n^2 units over a real field F and let B have type Rn with respect to F. By this we mean to assume that B contains a quantity a of grade n with respect to F, with minimum equation

$$\phi(\xi) \equiv \xi^n + \epsilon_1\xi^{n-1} + \ldots + \epsilon_n = 0 \quad (\epsilon_1, \ldots, \epsilon_n \text{ in } F), \tag{5}$$

and with polynomials

$$\theta_j(a), \quad \theta_1(a) = a \quad (j = 1, \ldots, n),$$

which are distinct and satisfy (5). A basis of B is then given by quantities

$$u_{jn+k} = a^{j-1}\gamma_k, \quad u_1 = y_1 = 1, \quad (j, k = 1, \ldots, n), \tag{6}$$

such that

$$y_j f(a) = f[\theta_j(a)]y_j \tag{7}$$

for every quantity $f(a)$ of the algebraic field $F(a)$. There exists integers $t_{j,k}$ such that

$$\theta_j[\theta_p(a)] = \theta_{t_{j,k}}(a)$$

and quantities $g_{j,k}(a)$ in $F(a)$ such that

$$y_k y_j = g_{j,k}\, y_{t_{j,k}}.$$

Let $\alpha_j = \theta_j(\alpha_1)$ be the ordinary complex roots of (5). The author has then proved that we may take $p = np'$,

$$A = \| A_{jk} \|, \; A_{j,j-1} = I(2p'), \; A_{jk} = 0(2p') \; (k \neq j-1),$$
$$A_{jn} = -\epsilon_{n+1-j}I(2p') \quad (j, k = 1, \ldots, n),$$

as a first representation of a. Let

$$y_j = \| y_{st}^{(j)} \| \quad (j, s, t = 1, \ldots, n)$$

be a first representation of y_j when a first representation of B exists and define matrices

$$\Lambda_j \equiv \Lambda_j(\alpha_y) \equiv \sum_{s=1}^{n} \alpha_1^{s-1} y_{s1}^{(j)}.$$

It can be shown that

$$\Lambda_s(\alpha_1)\; \Lambda_j(\alpha_s) = g_{j,s}\,(\alpha_1)\, \Lambda_{t_{j,s}}(\alpha_1).$$

Moreover, there exists a p-rowed second representation B_2 of B if and only if $p = np'$, and when this is true we can represent a by

$$\alpha = \| \alpha_{jk} \|, \; \alpha_{jk} = 0(p'), \; \alpha_{jj} = \alpha_j I(p'), \; (j \neq k; j, k = 1, \ldots, n),$$

and the y_j by

$$\eta_j = \| \eta_{st}^{(j)} \|, \; \eta_{st}^{(j)} = 0(p'), \; (t \neq t_{j,s}),$$
$$\eta_{st_{j,s}}^{(j)} = g_{j,s}(\alpha_1)I(p') \quad (j, s, t = 1, \ldots, n).$$

For the case where the α_j are all real and there exists a first representation of B we have

THEOREM 5. *A matrix ω with p rows and $2p$ columns of complex elements is a Riemann matrix having B as its multiplication algebra if and only if*

(i) *The matrix ω is isomorphic to a matrix*

$$\omega = \| \tau \wedge_j \alpha_j^{k-1} \| \quad (j, k = 1, \ldots, n).$$

(ii) *The p'-rowed and $2p'$-columned matrix τ is a Riemann matrix over $F(\alpha_1)$ and has a principal matrix $\Gamma(\alpha_1)$ such that there exist positive numbers $\beta_j(\alpha_1)$ in $F(\alpha_1)$ with the property that $\wedge_j \Gamma(\alpha_j) \wedge'_j = \beta_j \Gamma(\alpha_1)$.*

(iii) *The only matrices $G(\alpha_1)$ for which $\tau G(\alpha_1) \tau' = 0$ and $\wedge_j G(\alpha_j) \wedge'_j = \beta_j G(\alpha_1)$ are multiples of $\Gamma(\alpha_1)$ by numbers of $F(\alpha_1)$.*

The other case considered is that where the roots of (5) satisfy the condition of II of THEOREM 2. In this case we have

$$\theta_{1+n'}[\theta_j(\alpha_1)] = \theta_j[\theta_{1+n'}(\alpha_1)] = \theta_j(\bar{\alpha}_1) = \overline{\theta_j(\alpha_1)}$$

and

$$g_{j,1+n'}(\alpha_1) \wedge_{j+n'} (\alpha_1) = \wedge_{1+n'}(\alpha_1) \wedge_j (\bar{\alpha}_1).$$

From this we prove

THEOREM 6. *A matrix ω with p-rows and $2p$ columns of complex elements is a Riemann matrix over F with B as its multiplication algebra if and only if*

(i) *The matrix is isomorphic to*

$$\| \tau \wedge_j \alpha_j^{k-1} \|.$$

(ii) *The $2p'$-rowed square matrix*

$$\Omega_1 = \left\| \frac{\tau}{\overline{\tau \wedge_{1+n'}}} \right\|$$

is an Omega matrix over $F(\alpha_1)$ and has a principal matrix $\Gamma(\alpha_1)$ such that

$$\wedge_j \Gamma(\alpha_j) \overline{\wedge}'_j = \beta_j \Gamma(\alpha_1)$$

for positive β_j of $F(\alpha_1)$.

(iii) *The only matrices $G(\alpha_1)$ for which*

$$\tau G(\tau \wedge_{1+n'})' = 0 \text{ and } \wedge_j G(\alpha_j) \overline{\wedge}'_j = \beta_j G(\alpha_1)$$

are multiples of $\Gamma(\alpha_1)$ by numbers of $F(\alpha_1)$.

The above Theorem 5 has been applied to the case $n = 2$ and actual examples of pure Riemann matrices with non-commutative multiplication algebras have been constructed. It has also been shown that there exist no pure Riemann matrices with B a normal division algebra in sixteen units over the field of all rational numbers.

* For the properties of Riemann matrices see Bulletin of the National Research Council, number 63, *Selected Topics in Algebraic Geometry*, chapters 15, 16, 17, 1928.

ON DIRECT PRODUCTS, CYCLIC DIVISION ALGEBRAS, AND PURE RIEMANN MATRICES

BY A. ADRIAN ALBERT

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated March 7, 1930

We shall use the definitions and theorems on simple algebras of L. E. Dickson's *Algebras and Their Arithmetics*. A simple algebra over a non-modular field F is called *normal* if the only quantities of the algebra A which are commutative with every quantity of A are multiples of the modulus of A by scalars of F. Every normal simple algebra is the direct product of a normal division algebra and a total matric algebra and conversely. We shall use the symbol $\times$ to indicate *direct product* and have

THEOREM 1. *The direct product of two normal simple algebras is a normal simple algebra.*

THEOREM 2. *If $A = B \times C$ where A and B are normal simple algebras then C is a normal simple algebra.*

THEOREM 3. *If $A = B \times C$ where A and B are total matric algebras then C is a total matric algebra.*

THEOREM 4. *The direct product A of two algebraic fields $F(a)$ and $F(b)$ of orders m and n, respectively, with respect to the general non-nodular field F, is a division algebra if and only if the field $F(\xi,\eta)$ has order mn for any scalar root ξ of the minimum equation of a and scalar root η of the minimum equation of b.*

With this as a criterion we have

THEOREM 5. *A direct product $A = B \times C$ of two division algebras B and C is a division algebra if and only if there exists no b in B and c in C such that $F(b) \times F(c)$ is not a division algebra.*

As a consequence of Theorems 4 and 5 we have

THEOREM 6. *Let $A = B \times C$ where B and C are division algebras of relatively prime orders. Then A is a division algebra.*

We shall now consider the theory of the representations of algebras. A linear set D of p-rowed square matrices with scalar elements is called a representation of an algebra B over F if D is an algebra over F equivalent to B. We consider the case where B is a normal division algebra and prove

THEOREM 7. *If B is a normal division algebra of order n^2 over F there exists a normal simple algebra B_1 of order n^2 over F such that $A = B \times B_1$ is a total matric algebra over F.*

THEOREM 8. *If $M = B \times C$ is a total matric algebra over F, and B and C are normal division algebras, then B and C have the same order.*

THEOREM 9. *A normal division algebra B of order n^2 over F has a rep-*

resentation as an algebra of m-rowed square matrices with elements in F if and only if n^2 divides m.

THEOREM 10. *The algebra B_1 of* Theorem 7 *is a normal division algebra.*

Consider a cyclic (Dickson) algebra A of order n^2 over F so that A contains a quantity x whose minimum equation with respect to F is cyclic with respect to F. Let $\theta(x)$ be the polynomial whose n distinct iteratives are roots of the minimum equation of x so that A contains a quantity y such that $y^r x = \theta^r(x) y^r$ $(r = 0,1, \ldots, n-1)$ and $y^n = \gamma$ in F. We call γ the γ of A.

THEOREM 11. *Every cyclic division algebra A of order n^2 over F where $n = p_1^{e_1}, \cdots p_i^{e_t}$ and the p_i are distinct primes is a direct product of t cyclic algebras B_i, of orders $p_i^{2e_i}$, respectively, each a division algebra with the same γ as A. Conversely every direct product A of t cyclic division algebras B_i of orders p_i, with the p_i distinct primes, is a cyclic division algebra whose γ may be taken as the γ of all the Bi.*

The following theorem known for $p = 2,3$ has been proved for any prime p.

THEOREM 12. *A cyclic algebra of order p^2 over F, p a prime, is a division algebra if and only if γ is not the norm of any polynomial in x.*

It is known that a cyclic algebra is a division algebra if γ^r is not a norm for $r < n$ when the algebra has order n^2. However this condition does not seem to be necessary. In this connection we have proved

THEOREM 13. *Let A be a cyclic division algebra of order n^2, where $n = p_1^{e_1}, p_2^{e_2} \ldots p_t^{e_t}$ and the p_1 are distinct primes. Let $r = p_1 p_2 \ldots p_t$. Then if γ^s is a norm, s is divisible by r.*

THEOREM 14. *Let A be a cyclic division algebra of order n^2 over F and let γ^p be a norm where p is a prime. Then n is a power of p.*

Finally when B is a cyclic division algebra of order p^2, p a prime, we may determine explicitly the structure of the division algebra B_1 of Theorem 7. We have

THEOREM 15. *Algebra B_1 of* Theorem 7 *is equivalent to B when B is a cyclic division algebra of order p^2, p a prime.*

By the use of Theorems 9 and 14 we may prove two new results of extreme importance in the theory of Riemann matrices. Using Theorem 9, we have

THEOREM 16. *The multiplication index h of a pure Riemann matrix of genus p is a divisor of 2p.*

It was previously known merely that h was at most $2p$. We also have from Theorem 14 and certain results of a previous paper* of the author that

THEOREM 17. *Let ω be a pure Riemann matrix over a real field F and let ω have a multiplication algebra which is a normal division algebra of order n^2 over F and which contains a quantity a of grade n with respect to F and such that there exist n distinct polynomials in a satisfying the minimum equation $\phi(\xi) = 0$ of a with respect to F. Let $\phi(\xi) = 0$ have a real root or all imaginary*

roots such that the substitution carrying each complex root to its complex conjugate is commutative with the galois group of $\phi(\xi) = 0$. *Then n is necessarily a power of* 2.

The restriction we have made on the algebra in the above theorem amounts to just a little more than choosing the algebra to be a known normal division algebra, the only case that can be considered in detail at present, due to the lack of knowledge of the structure of division algebras.

* This number of these PROCEEDINGS, *On the Structure of Pure Riemann Matrices, etc.*

THE NEAR INFRA-RED ABSORPTION SPECTRUM OF CALCITE

BY JOSEPH W. ELLIS

PHYSICS DEPARTMENT, UNIVERSITY OF CALIFORNIA AT LOS ANGELES

Communicated February 18, 1930

An investigation of the infra-red absorption spectrum of calcite below 3μ was undertaken by the writer about a year ago, but the work was interrupted by the removal of this laboratory to a new campus. From the work of Schaefer, Bormuth and Matossi[1] on the absorption spectra of calcite and other carbonates it seemed certain that if sections of crystals thicker than those employed by these investigators were used, new bands in the higher frequency region should be found. In the meantime Plyler[2] has announced his finding of four such bands at 2.20, 2.00, 1.90 and 1.76μ. The completion of the present investigation has revealed several additional features of interest which did not show up in Plyler's study. These new features, which will be discussed in order, include:

(1) Three unreported bands with wave-lengths shorter than 1.7μ.

(2) A doublet structure in practically all of the bands in this short wave region.

(3) An almost complete disappearance of all of the bands when the light is nearly completely plane polarized at right angles to the optic axis.

(4) An increase with amplitude in the force coefficient associated with one of the four independent modes of vibration of the carbonate ion.

The three new bands which were found were quite distinct although relatively weak. However, it is probable that they would not have been observed without the use of a registering spectrograph. To get the proper thickness of crystal to reveal them, three calcite rhombs were arranged in tandem, making a total path of 7.0 cm. The light rays, without any preliminary polarization, traversed principal sections along paths oblique to the optic axis. The ordinary image of the lamp source, which was well separated from the extraordinary image, was focused upon the slit.

The bands, like the stronger ones of greater wave-length, were not sharp but indicated a double structure. Because of this weakness, however, only the most intense parts of the bands, probably the stronger components of doublets, were measured. These wave-length values are 1.65, 1.56 and 1.47μ. Of the three, the 1.56μ band is most intense.

The following bands, which are not new, have for the first time been clearly recorded as doublets, the longer wave component being in every instance much more intense than its companion: 2.58, 2.52; 2.36, 2.31; 2.17, 2.12; 2.01, 1.97; 1.87, 1.84; and 1.75, 1.72μ. The order of accuracy in these determinations is $\pm 0.01\mu$. Plyler gives for the centers of these unresolved bands: 2.52, 2.34, 2.20, 2.00, 1.90 and 1.76μ. The discrepancies in some instances far exceed the degree of accuracy in the present measurements.

To discuss the behavior of these bands in polarized light it will be necessary to summarize briefly the findings of previous investigations upon calcite. The similarity of the absorption spectra of all crystalline carbonates has led to the conclusion, now generally accepted, that all of the bands in the near infra-red are produced by vibrations within the carbonate group. Again, the analysis of calcite, i.e., calcium carbonate, by means of x-rays has definitely established the fact that in this crystal the atoms constituting the carbonate ion lie in a plane, the three oxygen atoms at the corners of an equilateral triangle with the carbon at the center. These triangular ions are themselves coplanar, making up exclusively every alternate 111 plane of the crystal lattice. Experimental measurements upon absorption and reflection spectra in the region 3–15μ, as well as certain theoretical considerations, have led investigators to the conclusion that three fundamental modes of vibration characterize the motions of the carbon and oxygen atoms in the plane of the ion, whereas there is but one mode of vibration at right angles to this plane. [For the references to the main contributions to the literature of this subject see the papers cited in (1) and (2) at the end of this article.] The approximate wave-length positions of bands associated with these vibrations are 14, 9 and 7μ, and 11μ, respectively. Of these the one near 9μ is "optically inactive" and shows up spectroscopically only in the Raman spectrum and through its entrance into combination with the other modes of vibration to form higher frequency "combination bands."

Now in the present investigation when the extraordinary image, isolated from the ordinary image after the light had traversed a rhomb of 2 cm. thickness, was focused upon the slit, all of the bands with wave-lengths greater than 1.7μ produced by the ordinary rays were recorded, but with diminished intensity. The fact that previous investigators have found that the frequency associated with the 11μ maximum does not enter into strong combination bands suggests that all of the bands below 3μ are

combinations or overtones of the frequencies associated with vibrations of the carbonate ion in its own plane. This notion was strongly supported by the appearance of all the bands with considerable intensity when ordinary rays were used and by their diminution in intensity when the extraordinary rays alone were used. For in the former case even without preliminary polarization of the light all of the vibrations were at right angles to the optic axis, i.e., in the plane of the ion; and in the latter case, although they were restricted to a plane including the optic axis, the vibrations still had a component in the plane of the ion.

To restrict the vibrations of the light to a definite plane within the crystal, light which was nearly completely plane polarized was sent through a specimen cut parallel to the optic axis. The polarization was produced by means of a pile of thirteen thin microscope cover glasses. Failure to produce a beam of light absolutely plane polarized was effected by the natural limitations on the part of such a pile of plates and by the fact that the slight convergence in the rays prevented them from all striking the plates exactly at the polarizing angle. However, the absorption bands which appeared with considerable intensity when the vibrations were nearly restricted to a plane at right angles to the optic axis appeared as mere traces when the vibrations were nearly restricted to the direction of the optic axis.

The author concurs in the opinion of Plyler that the bands near 2.34 and 1.75μ are the second and third overtones, respectively, of the 7μ fundamental. But it must be pointed out that the frequencies of these two bands taken with those of the first overtone and fundamental cannot be fitted to the anharmonic type of formula which has found extensive applications in the spectra of gases and liquids. The Raman effect at the hands of Schaefer, Matossi and Aderhold[3] has probably yielded the best value for the wave-length of the 7μ maximum. The single value of 6.92μ which these investigators found is interpreted here as marking the location of the longer wave component of this fundamental band. Absorption measurements have not been very helpful in revealing the exact nature of the absorption near 7μ because of the great intensity of absorption in that region. The best determination of the value of the first overtone doublet is found in the absorption measurements of Schaefer, Bormuth and Matossi.[1] These authors give 3.47 and 3.35μ for these two components, the former being the more intense. Taking the components of longer wave-length only, the following table of products can be constructed:

TABLE 1

$1 \times 6.92 = 6.92$	$3 \times 2.36 = 7.08$
$2 \times 3.47 = 6.94$	$4 \times 1.75 = 7.00$

If the bands obeyed the usual anharmonic formula the values for these products should increase in a regular manner. It is clearly seen that the wave-length of the third overtone is too small, i.e., the frequency of vibration associated with it is too large. Although, so far as the writer is aware, no case of a similar non-harmonic series is on record, this example should elicit no surprise. For, whatever may be the nature of the motions of the carbon and oxygen nuclei associated with this series of frequencies, when amplitudes of sufficient magnitude are reached, it would seem that rather large repulsive forces arising from surrounding ions might be introduced. This would produce an increase in the frequencies of vibration.

Table 2, which is merely an extension to and a correction of a table originated by Schaefer, Bormuth and Matossi[1] and continued by Plyler,[2] shows the correlation of the bands associated with vibrations in the plane of the ion. It will be observed that the three new absorption maxima find their places quantitatively in the system of combination frequencies which leaves no band unaccounted for and which requires no band not observed. The values for these three fundamental frequencies have been taken from the Raman spectra recorded by Schaefer, Matossi and Aderhold.[3] The values for the λ calc. and ν calc. columns have been worked out on the basis of the more intense longer wave component of the doublets. There is only one discrepancy between the calculated values and those observed in this investigation, namely, in the 2.58μ band. The location of this band almost exactly half way between the well-known 2.49μ benzene band and the 2.66μ water-vapor band allows it to be determined with considerable accuracy. The only way in which to account for the discrepancy is to assume that the value given by Schaefer, Bormuth and Matossi for the first overtone band is too low. A slightly higher value for this wave-length would also improve the agreement in table 1.

The doublet nature of the overtone and combination bands is of interest. The combination bands apparently derive their doubleness from that of the overtone bands which enter into the combination. Now Schaefer, Bormuth and Matossi, following Brester, have suggested that the doubleness of the 7μ, as well as of the 14μ, band may arise from a slight departure of the carbonate ion from a true equilateral triangular form, and that the frequency of vibration of the type which produces absorption near 7μ will depend upon the orientation of the plane of polarization of the light rays with respect to this less symmetrical triangular ion. To seek a possible experimental confirmation of this interpretation the absorption by a 1 mm. crystal section cut perpendicularly to the optic axis was studied, plane polarized light being used. Six records of the overtone doublet at 2.36, 2.31μ and of the combination doublet at 2.58, 2.52μ were obtained, the conditions of the experiment differing only in that before each new record was taken the crystal was rotated approximately 30°

about the light beam, i.e., about the optic axis, as an axis. No variation in the intensity ratios between the components of the doublets was observed. It should be emphasized that the registering spectrograph affords a very sensitive means for observing such a variation, should it exist. From this failure to detect a variation in the intensity ratios it

TABLE 2

DESIGNATION	λ OBS.	λ CALC.	ν OBS.	ν CALC.	OBSERVERS
ν_2	13.94μ	....	717 cm.$^{-1}$	....	S, M, A*
ν_0	9.18	....	1089	....	S, M, A
ν_1	6.92	....	1445	....	S, M, A
(0) + ν_2 + ν_0	5.55	5.54μ	1800	1806 cm.$^{-1}$	S, B, M**
ν_1 + ν_2	4.64	4.62	2155	2162	S, B, M
ν_1 + ν_0	3.93	3.94	2545	2534	S, B, M
$2\nu_1$	3.47 3.35	....	2880 2990	....	S, B, M
ν_1 + ν_2 + ν_0	3.10	3.08	3230	3251	S, B, M
$2\nu_1$ + ν_2	2.79	2.78	3580	3597	S, B, M
$2\nu_1$ + ν_0	2.58 2.52	2.52	3880 3970	3969	Ellis
$3\nu_1$	2.36 2.31	....	4240 4330	....	Ellis
$2\nu_1$ + ν_2 + ν_0	2.17 2.12	2.17	4610 4720	4606	Ellis
$3\nu_1$ + ν_2	2.01 1.97	2.01	4970 5080	4960	Ellis
$3\nu_1$ + ν_0	1.87 1.84	1.87	5350 5440	5329	Ellis
$4\nu_1$	1.75 1.72	....	5720 5820	....	Ellis
$3\nu_1$ + ν_2 + ν_0	1.65	1.65	6060	6046	Ellis
$4\nu_1$ + ν_2	1.56	1.55	6410	6437	Ellis
$4\nu_1$ + ν_0	1.47	1.47	6800	6809	Ellis

* S, M, A—Schaefer, Matossi, Aderhold, by Raman effect.
** S, B, M—Schaefer, Bormuth, Matossi, by absorption.

cannot necessarily be concluded that the triangular ion is not unsymmetrical for there still exists the possibility that such unsymmetrical groups may be differently oriented in the crystal. But it can be concluded that the doubleness in this instance is not completely analogous to the doubleness exhibited by biaxial crystals, for in these latter instances intensity variations can be secured by changing the plane of polarization with respect to the crystalline axes.

Within the limits of experimental certainty the frequency separations of the members of the non-harmonic series originating in the 7μ band are constant. If they are constant and if an attempt is made to formulate equations to relate the long and short wave components independently then it must be recognized that one or both of the resulting polynomial equations must include a constant term. A short paper dealing with the appearance of constant terms in the equations relating the vibrational doublets in another instance has been submitted elsewhere for publication.

[1] Schaefer, Bormuth and Matossi, *Zeits. Physik*, **39**, 648 (1926).

[2] E. K. Plyler, *Phys. Rev.*, **33**, 948 (1929).

[3] Schaefer, Matossi and Aderhold, *Phys. Zeits.*, **30**, 581 (1929).

THE EFFECT OF THE ANNIHILATION OF MATTER ON THE WAVE-LENGTH OF LIGHT FROM THE NEBULAE

BY RICHARD C. TOLMAN

NORMAN BRIDGE LABORATORY OF PHYSICS, PASADENA, CALIFORNIA

Communicated March 12, 1930

§ 1. *Introduction.*—On the basis of the general theory of relativity, Einstein[1] and de Sitter[2] have each proposed a line element which might correspond to the large-scale metrical properties of the universe as a whole, neglecting local deviations due to the neighborhood of stars or stellar systems. Neither of these line elements, however, has proved entirely satisfactory.

The Einstein line element, which can be written in the form

$$ds^2 = -\frac{dr^2}{1-\frac{r^2}{R^2}} - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 + dt^2, \qquad (1)$$

has a reasonably satisfactory derivation, since it can be obtained from the equations of general relativity, if we assume that the universe can be regarded on a large scale as though permanently filled with a homogeneous fluid, and assume that the cosmological constant Λ, which occurs in the relativity equations connecting metric and the distribution of matter, can be regarded as a parameter determined by the density and pressure of this fluid.[3] This line element, however, would not lead to any red shift in the light coming from distant objects, as is actually found in light from the extra-galactic nebulae.[4]

The de Sitter line element, which can be written in the form

$$ds^2 = -\frac{dr^2}{1-\frac{r^2}{R^2}} - r^2 d\theta^2 - r^2 \sin^2\theta d\phi^2 + \left(1-\frac{r^2}{R^2}\right) dt^2, \quad (2)$$

can be derived from the equations of general relativity if we assume a completely empty universe, and assume that the cosmological constant Λ is a new fundamental constant connected with a natural curvature of free space. The theoretical basis for this line element is hence not entirely pleasing, since the actual presence of matter in the universe must be regarded as producing a disturbance away from the proposed form of line element, instead of actually determining that form as in the case of the Einstein line element. Furthermore, the actual red-shift in the light from the nebulae can be accounted for on the basis of the above equation only if we ascribe to these nebulae just the right velocities of recession to account for the observed facts. This seems an *ad hoc* method of accounting for the observations, but its arbitrary character appears somewhat diminished by the considerations of Robertson,[5] who has shown by a very clever transformation of coördinates that particles having the right velocities of recession to account for the linear increase in shift with distance would all be equivalent to each other, in the sense that observers located on different particles would obtain the same results for the relation between red shift and distance.

Since neither of these previously proposed line elements is completely satisfactory, it is next natural to inquire if some other form of static line element could account for the red shift in a more natural manner. I have already investigated this problem,[6] however, and found that the line elements of Einstein and de Sitter, including the special relativity case obtained with $R = \infty$, are the only static ones which would agree with those restrictions as to spherical symmetry, symmetry with respect to past and future time, uniform density of material throughout the universe, etc., which it seems natural to impose; and this conclusion has since been confirmed by Robertson.[7]

Thus it appears probable that no static line element will successfully account for the red shift from distant objects, and if this phenomenon is actually due to the metric for the universe instead of to some more immediate cause such as that suggested by Zwicky,[8] we must turn to the consideration of *non-static* line elements. Such a change in attack, however, is not an unnatural one to make since, as I pointed out in the article already mentioned,[6] static line elements take no cognizance of any universal evolutionary process which may be going on. Hence the actual line element may well be non-static if the matter in the universe is not in a steady state.

As the basic hypothesis for the present article we shall assume, in

agreement with the contemporary opinion of astrophysicists, that the matter in the universe is not in a steady state but rather that a general transformation of matter into radiant energy is taking place throughout the universe at the rate necessary to account for the radiation from stellar objects. If such a process is going on, the line element for the universe cannot be static, but necessarily must be non-static, since matter and the radiation produced from it would not have the same effect on the gravitational field.

In the following sections, we shall attempt to deduce the form of line element which would correspond to the transformation of matter into radiation postulated above, and by introducing reasonable simplifications and assumptions, shall actually be able to obtain as a first approximation for the form of the line element the expression

$$ds^2 = -\frac{e^{2kt}}{\left(1+\frac{4r^2}{R^2}\right)^2}(dx^2+dy^2+dz^2)+dt^2, \qquad (3)$$

where R is a constant having approximately the same significance as in the Einstein line element and k a constant directly related, on the one hand, to the average rate of transformation of matter into radiation and, on the other hand, to the known shift in wave-length with distance. Furthermore by comparison with observational data, we shall show that the value of k necessary to account for the empirical relation between red shift and distance for the extra galactic nebulae found by Hubble and Humason,[4] falls close to the range of values which would correspond to the rates of transformation of matter into radiation for actual stars of different types.

§ 2. *The General Form of the Line Element.*—To determine the form of the line element, we shall begin by ascribing spatial spherical symmetry to it, in order to agree with the idea that matter and radiation are uniformly distributed on the average throughout the universe. And we shall also ascribe to it symmetry with respect to past and future time in order to agree with the principle of dynamical reversibility. We can then write the line element in the general form

$$ds^2 = -e^{\mu}(dx^2+dy^2+dz^2)+e^{\nu}dt^2, \qquad (4)$$

where μ and ν are as yet undetermined functions of the distance from the origin $r = \sqrt{x^2+y^2+z^2}$ and of the time t.

As our next restriction on the form of the line element, we shall require that particles which are stationary in the chosen coördinates x, y, z shall remain permanently stationary and not be subject to accelerations. Without the possibility of imposing this restriction, it would be difficult to construct a stable model of the universe. For the acceleration of a

particle, however, we have in accordance with the theory of relativity the well-known equation for a geodesic

$$\frac{d^2x_\alpha}{ds^2} + \{\rho\sigma,\alpha\}\frac{dx_\rho}{ds}\frac{dx_\sigma}{ds} = 0. \tag{5}$$

And considering for example accelerations in the x-direction, for particles having zero spatial velocities

$$\frac{dx}{ds} = \frac{dy}{ds} = \frac{dz}{ds} = 0. \tag{6}$$

equation (5) reduces to

$$\frac{d^2x}{ds^2} + \{44,1\}\left(\frac{dt}{ds}\right)^2 = 0. \tag{7}$$

Thus, if the acceleration d^2x/ds^2 is to be equal to zero, we are led to the conclusion that the line element must be such as to make the Christoffel 3-index symbol $\{44,1\}$ itself equal to zero. For the line element (4), however, we can easily calculate the value of this symbol (see equation 18 below), and obtain

$$\{44,1\} = -\frac{1}{2}g^{11}\frac{\partial g_{44}}{\partial x} = \frac{1}{2}e^{\mu+\nu}\frac{\partial \nu}{\partial x} = 0. \tag{8}$$

The result shows that ν must be independent of x and hence by symmetry independent of y and z as well.

Hence ν is a quantity which must be a function of the time t alone, and by introducing a new time variable, t', defined by the equation

$$e^{\nu(t)}dt^2 = dt'^2, \tag{9}$$

and then dropping the primes we can necessarily reduce the line element (4) to the form

$$ds^2 = -e^{\mu}(dx^2 + dy^2 + dz^2) + dt^2, \tag{10}$$

where μ still remains as an undetermined function of r and t.

Nevertheless, a restriction on the form of μ can now easily be obtained by the following consideration. For the proper spatial volume dV_0, contained between non-accelerated particles located on the boundary of the volume element determined by dx, dy, dz, we can evidently write

$$dV_0 = e^{3\mu/2}\,dx\,dy\,dz, \tag{11}$$

and the logarithmic differential of this with respect to the time will evidently be

$$\frac{\partial \log dV_0}{\partial t} = \frac{3}{2}\frac{\partial \mu}{\partial t}. \tag{12}$$

Since, however, in accordance with the line element (10) the proper time is the same for all stationary observers throughout the universe, we can maintain our original assumption of uniform conditions throughout the universe, only if the percentage change with the time in the proper volume enclosed by "stationary" particles is independent of position. We must hence have

$$\frac{\partial}{\partial r}\frac{\partial \log dV_0}{\partial t} = \frac{3}{2}\frac{\partial^2 \mu}{\partial r \partial t} = 0, \tag{13}$$

and can write

$$\frac{\partial^2 \mu}{\partial r\, \partial t} = 0 \qquad \text{or} \qquad \mu = f(r) + g(t), \tag{14}$$

where f and g are undetermined functions of r and t, respectively.

Summarizing the present status of our information as to the general form of the line element, we may now write

$$ds^2 = -e^{\mu}(dx^2 + dy^2 + dz^2) + dt^2$$

with

$$\begin{aligned} g_{11} &= g_{22} = g_{33} = -e^{\mu} & g_{44} &= 1 \\ g^{11} &= g^{22} = g^{33} = -e^{-\mu} & g^{44} &= 1 \\ g_{\rho\sigma} &= g^{\rho\sigma} = 0 \qquad (\rho \neq \sigma) & \sqrt{-g} &= e^{3\mu/2} \end{aligned} \tag{15}$$

and

$$\mu = f(r) + g(t)$$
$$r = \sqrt{x^2 + y^2 + z^2}.$$

§3. *The Christoffel Symbols.*—To proceed further in our determination of the form of the line element, we must next calculate the values of the Christoffel 3-index symbols corresponding to the line element given. These are defined in accordance with the equation

$$\left\{\mu\nu,\sigma\right\} = \frac{1}{2}\, g^{\sigma\lambda}\left(\frac{\partial g_{\mu\lambda}}{\partial x_{\nu}} + \frac{\partial g_{\nu\lambda}}{\partial x_{\mu}} - \frac{\partial g_{\mu\nu}}{\partial x_{\lambda}}\right), \tag{16}$$

which for our line element evidently reduces to

$$\left\{\mu\nu,\sigma\right\} = \frac{1}{2}\, g^{\sigma\sigma}\left(\frac{\partial g_{\mu\sigma}}{\partial x_{\nu}} + \frac{\partial g_{\nu\sigma}}{\partial x_{\mu}} - \frac{\partial g_{\mu\nu}}{\partial x_{\sigma}}\right) \quad \text{(not summed)}, \tag{17}$$

and taking μ, ν and σ as *different* indices we obtain the four following cases.

$$\begin{aligned}
\{\mu\mu,\mu\} &= \frac{1}{2} g^{\mu\mu} \frac{\partial g^{\mu\mu}}{\partial x_\mu} = \frac{1}{2} \frac{\partial \log g_{\mu\mu}}{\partial x_\mu} \\
\{\mu\mu,\nu\} &= -\frac{1}{2} g^{\nu\nu} \frac{\partial g_{\mu\mu}}{\partial x_\nu} \\
\{\mu\nu,\nu\} &= \{\nu\mu,\nu\} = \frac{1}{2} g^{\nu\nu} \frac{\partial g_{\nu\nu}}{\partial x_\mu} = \frac{1}{2} \frac{\partial \log g_{\nu\nu}}{\partial x_\mu} \\
\{\mu\nu,\sigma\} &= 0.
\end{aligned} \tag{18}$$

Applying these equations, we can then easily calculate as the only non-vanishing Christoffel symbols, corresponding to the line element defined by (15)

$$\begin{aligned}
\{11,1\} &= \frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{22,2\} = \frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{33,3\} = \frac{1}{2}\frac{\partial \mu}{\partial z} \quad \{41,1\} = \frac{1}{2}\frac{\partial \mu}{\partial t} \\
\{11,2\} &= -\frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{22,1\} = -\frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{33,1\} = -\frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{42,2\} = \frac{1}{2}\frac{\partial \mu}{\partial t} \\
\{11,3\} &= -\frac{1}{2}\frac{\partial \mu}{\partial z} \quad \{22,3\} = -\frac{1}{2}\frac{\partial \mu}{\partial z} \quad \{33,2\} = -\frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{43,3\} = \frac{1}{2}\frac{\partial \mu}{\partial t} \\
\{11,4\} &= \frac{1}{2} e^{\mu} \frac{\partial \mu}{\partial t} \quad \{22,4\} = \frac{1}{2} e^{\mu} \frac{\partial \mu}{\partial t} \quad \{33,4\} = \frac{1}{2} e^{\mu} \frac{\partial \mu}{\partial t} \\
\{12,2\} &= \frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{21,1\} = \frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{31,1\} = \frac{1}{2}\frac{\partial \mu}{\partial z} \\
\{13,3\} &= \frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{23,3\} = \frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{32,2\} = \frac{1}{2}\frac{\partial \mu}{\partial z} \\
\{12,1\} &= \frac{1}{2}\frac{\partial \mu}{\partial y} \quad \{21,2\} = \frac{1}{2}\frac{\partial \mu}{\partial x} \quad \{31,3\} = \frac{1}{2}\frac{\partial \mu}{\partial x} \\
\{13,1\} &= \frac{1}{2}\frac{\partial \mu}{\partial z} \quad \{23,2\} = \frac{1}{2}\frac{\partial \mu}{\partial z} \quad \{32,3\} = \frac{1}{2}\frac{\partial \mu}{\partial y} \\
\{14,1\} &= \frac{1}{2}\frac{\partial \mu}{\partial t} \quad \{24,2\} = \frac{1}{2}\frac{\partial \mu}{\partial t} \quad \{34,3\} = \frac{1}{2}\frac{\partial \mu}{\partial t}
\end{aligned} \tag{19}$$

§ 4. *The Contracted Riemann-Christoffel Tensor.*—Using these values of the Christoffel symbols, we can now calculate the components of the contracted Riemann-Christoffel tensor, which are given by the well-known equation

$$G_{\mu\nu} = -\frac{\partial}{\partial x_\alpha}\{\mu\nu,\alpha\} + \{\mu\alpha,\beta\}\{\nu\beta,\alpha\} + \frac{\partial^2}{\partial x_\mu \partial x_\nu} \log \sqrt{-g} - \{\mu\nu,\alpha\} \frac{\partial}{\partial x_\alpha} \log \sqrt{-g}. \tag{20}$$

As a result of the calculation, noting the restriction on the form of μ given

by equation (14), we obtain as the only non-vanishing components of this tensor

$$G_{11} = \frac{\partial^2\mu}{\partial x^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial y^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial z^2} + \frac{1}{4}\left(\frac{\partial\mu}{\partial y}\right)^2 + \frac{1}{4}\left(\frac{\partial\mu}{\partial z}\right)^2 - \frac{1}{2}e^{\mu}\frac{\partial^2\mu}{\partial t^2} - \frac{3}{4}e^{\mu}\left(\frac{\partial\mu}{\partial t}\right)^2$$

$$G_{22} = \frac{\partial^2\mu}{\partial y^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial x^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial z^2} + \frac{1}{4}\left(\frac{\partial\mu}{\partial x}\right)^2 + \frac{1}{4}\left(\frac{\partial\mu}{\partial z}\right)^2 - \frac{1}{2}e^{\mu}\frac{\partial^2\mu}{\partial t^2} - \frac{3}{4}e^{\mu}\left(\frac{\partial\mu}{\partial t}\right)^2$$

$$G_{33} = \frac{\partial^2\mu}{\partial z^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial x^2} + \frac{1}{2}\frac{\partial^2\mu}{\partial y^2} + \frac{1}{4}\left(\frac{\partial\mu}{\partial x}\right)^2 + \frac{1}{4}\left(\frac{\partial\mu}{\partial y}\right)^2 - \frac{1}{2}e^{\mu}\frac{\partial^2\mu}{\partial t^2} - \frac{3}{4}e^{\mu}\left(\frac{\partial\mu}{\partial t}\right)^2$$

$$G_{44} = \frac{3}{2}\frac{\partial^2\mu}{\partial t^2} + \frac{3}{4}\left(\frac{\partial\mu}{\partial t}\right)^2 \qquad (21)$$

$$G_{12} = G_{21} = \frac{1}{2}\frac{\partial^2\mu}{\partial x\partial y} - \frac{1}{4}\frac{\partial\mu}{\partial x}\frac{\partial\mu}{\partial y}$$

$$G_{13} = G_{31} = \frac{1}{2}\frac{\partial^2\mu}{\partial x\partial z} - \frac{1}{4}\frac{\partial\mu}{\partial x}\frac{\partial\mu}{\partial z}$$

$$G_{23} = G_{32} = \frac{1}{2}\frac{\partial^2\mu}{\partial y\partial z} - \frac{1}{4}\frac{\partial\mu}{\partial y}\frac{\partial\mu}{\partial z}$$

In addition we can calculate from the above results for the invariant spur of this tensor

$$G = g^{\mu\nu}G_{\mu\nu} = e^{-\mu}\left[2\left\{\frac{\partial^2\mu}{\partial x^2} + \frac{\partial^2\mu}{\partial y^2} + \frac{\partial^2\mu}{\partial z^2}\right\} + \frac{1}{2}\left\{\left(\frac{\partial\mu}{\partial x}\right)^2 + \left(\frac{\partial\mu}{\partial y}\right)^2 + \left(\frac{\partial\mu}{\partial z}\right)^2\right\}\right] + 3\frac{\partial^2\mu}{\partial t^2} + 3\left(\frac{\partial\mu}{\partial t}\right)^2 \qquad (22)$$

and introducing the variable $r = \sqrt{x^2 + y^2 + z^2}$, this can be rewritten as

$$G = -e^{-\mu}\left[2\frac{\partial^2\mu}{\partial r^2} + \frac{4}{r}\frac{\partial\mu}{\partial r} + \frac{1}{2}\left(\frac{\partial\mu}{\partial r}\right)^2\right] + 3\frac{\partial^2\mu}{\partial t^2} + 3\left(\frac{\partial\mu}{\partial t}\right)^2. \qquad (23)$$

§ 5. *The Dependence of the Line Element on the Radius.*—We have now obtained sufficient material so that we can return to the problem of determining the form of the line element. In accordance with the fundamental equation of general relativity connecting the metric with the distribution of matter and energy, the invariant G is connected with the proper density of matter ρ_0 by the equation

$$8\pi\rho_0 = G - 4\Lambda, \qquad (24)$$

where Λ is the cosmological constant. We have assumed from the start, however, that the average properties of the universe are independent of position, and on account of the large scale of our interests are neglecting the local differences due to the neighborhood of individual stars or stellar systems. Hence we must take ρ_0 as independent of r, and since Λ is

in any case a constant we must have G also independent of r. Returning then to the expression for G given by equation (23), and remembering in accordance with equation (14) that $\partial\mu/\partial t$ is independent of r, we can write

$$e^{-\mu}\left[2\frac{\partial^2\mu}{\partial r^2}+\frac{4}{r}\frac{\partial\mu}{\partial r}+\frac{1}{2}\left(\frac{\partial\mu}{\partial r}\right)^2\right]+A=0, \tag{25}$$

where A is a quantity which is independent of r.

We have thus obtained a second-order differential equation to be solved for μ as a function of r. A first integral of this equation can be found in the following form,[9] as can be verified by resubstitution into equation (24),

$$\frac{\partial\mu}{\partial r}=-\frac{2}{r}\left(1\pm\sqrt{1-\frac{Ae^{\mu}r^2}{6}+\frac{C_1e^{-\mu/2}}{A^{1/2}r}}\right)$$

where C_1 is an arbitrary constant of integration. For our case, nevertheless, this constant must evidently have the value zero and the minus sign must be chosen on the right-hand side of the equation, since otherwise the quantity $\partial\mu/\partial r$ would become infinite at the origin. Making this choice, however, we find that $\partial\mu/\partial r$ approaches zero instead of infinity at the origin, in satisfactory agreement with our experimental knowledge that the special theory of relativity is approximately valid for a limited region in space and time.

Giving C_1 then the value zero, and using the minus sign, the equation reduces to

$$\frac{\partial\mu}{\partial r}=-\frac{2}{r}+\frac{2}{r}\sqrt{1-\frac{Ae^{\mu}r^2}{6}}, \tag{26}$$

and this can be integrated to give us

$$e^{\mu}=\frac{24C_2}{A(1+C_2r^2)^2}, \tag{27}$$

where C_2 is the arbitrary constant of integration. This final result can easily be verified by substitution back into equation (26) or better yet by substitution into the original equation (25) itself.

We have thus obtained the desired information as to the dependence of the line element on the radius r. And noting in accordance with equation (14) that e^{μ} can in any case be taken as the product of a function of r and t, we can now write

$$e^{\mu}=\frac{e^{g(t)}}{\left(1+\frac{r^2}{4R^2}\right)^2} \tag{28}$$

or for the line element as a whole

$$ds^2 = -\frac{e^{g(t)}}{\left(1+\frac{r^2}{4R^2}\right)^2}(dx^2+dy^2+dz^2)+dt^2, \qquad (29)$$

where we have substituted $1/4R^2$ for the constant C_2, in order to facilitate the comparison of this expression with that for the Einstein line element.

§ 6. *The Distribution of Matter and Radiation in the Universe.*—We may now investigate the distribution of matter and radiation in the universe which would correspond to this line element. Substituting the value for e^{μ} given by equation (28) into equations (21) for the Riemann-Christoffel tensor and raising suffixes for later convenience, we can now calculate as the *only* non-vanishing components of this tensor

$$G_1^1 = G_2^2 = G_3^3 = \frac{2}{R^2}e^{-g} + \frac{1}{2}\frac{d^2g}{dt^2} + \frac{3}{4}\left(\frac{dg}{dt}\right)^2$$

$$G_4^4 = \frac{3}{2}\frac{d^2g}{dt^2} + \frac{3}{4}\left(\frac{dg}{dt}\right)^2 \qquad (30)$$

with

$$G = \frac{6}{R^2}e^{-g} + 3\frac{d^2g}{dt^2} + 3\left(\frac{dg}{dt}\right)^2.$$

These quantities, however, are related to the energy-momentum tensor T_{μ}^{ν} in accordance with the fundamental equations

$$-8\pi T_{\mu}^{\nu} = G_{\mu}^{\nu} - \frac{1}{2}Gg_{\mu}^{\nu} + \Lambda g_{\mu}^{\nu}. \qquad (31)$$

Moreover, in applying this equation to cosmological considerations, our scale of interest is so large, that we shall regard our simplified model of the universe as though filled with a perfect fluid, and can then use for the energy-momentum tensor the well-known expression for such a fluid

$$T^{\mu\nu} = (\rho_{00} + p_0)\frac{dx_{\mu}}{ds}\frac{dx_{\nu}}{ds} - g^{\mu\nu}p_0, \qquad (32)$$

where ρ_{00} is the proper macroscopic density of the fluid, p_0 its proper pressure, and the quantities dx_{μ}/ds are macroscopic "velocities." Furthermore, in order to construct a stable model of the universe we have already imposed, in § 2, the condition that particles which are stationary in our coördinates shall not be subject to acceleration, and hence can take all the "velocities" dx_{μ}/ds as zero, except for the case $\mu = 4$. Under these circumstances, using our line element, and lowering suffixes we easily find as the only non-vanishing components for the energy-momentum tensor

$$T_1^1 = T_2^2 = T_3^3 = -p_0 \qquad T_4^4 = p_{00}. \tag{33}$$

Combining equations (30), (31) and (33), we now readily obtain for the proper pressure p_0, proper macroscopic density ρ_{00}, and proper density of matter ρ_0 (see equation 24)

$$\begin{aligned} 8\pi p_0 &= -\frac{1}{R^2}e^{-g} - \frac{d^2g}{dt^2} - \frac{3}{4}\left(\frac{dg}{dt}\right)^2 + \Lambda \\ 8\pi\rho_{00} &= \frac{3}{R^2}e^{-g} + \frac{3}{4}\left(\frac{dg}{dt}\right)^2 - \Lambda \\ 8\pi\rho_0 &= \frac{6}{R^2}e^{-g} + 3\frac{d^2g}{dt^2} + 3\left(\frac{dg}{dt}\right)^2 - 4\Lambda. \end{aligned} \tag{34}$$

§ 7. *The Dependence of the Line Element on Time.*—We must now return to the still unsolved problem of the dependence of the line element on the time, i.e., the dependence of g on t. With an exact knowledge of the changes taking place in pressure and density, we could theoretically use equations (34) to determine the exact dependence of g on t. In the absence of exact knowledge, however, we can start by assuming some simple empirical formula, as expressing the dependence of g on t, and then use what meager information we do have to evaluate as many constants as possible in this empirical formula.

As a simple empirical formula for the dependence of g on t, the following naturally suggests itself

$$g = a + bt + ct^2 + \ldots \tag{35}$$

as giving the value of g in the neighborhood of the time $t = 0$.

From the form of the line element (29), it is evident that the first constant in this formula a merely depends on the choice of units for laying off the system of spatial coördinates, and hence would naturally be taken as zero in order that the line element should reduce to the special relativity form at $t = 0$ and in the neighborhood of the origin $r = 0$. For the second constant b, it will be better to use the symbol $2k$ in order to avoid fractions in the later calculations. As for the higher terms, we do not as yet have sufficient information for their evaluation and we shall have to omit them, introducing at this point the assumption that, in the neighborhood of the time $t = 0$, the quantity g can be given with sufficient approximation by a formula which is linear in t.

We can then write as our chosen empirical expression for the dependence of g on t,

$$g = 2kt \qquad \frac{dg}{dt} = 2k \qquad \frac{d^2g}{dt^2} = 0 \tag{36}$$

and for the line element as a whole

$$ds^2 = -\frac{e^{2kt}}{\left(1+\frac{r^2}{4R^2}\right)^2}(dx^2+dy^2+dz^2)+dt^2. \qquad (37)$$

§ 8. *Relation of the Quantity k to the Rate of Transformation of Matter into Radiation.*—With the help of our explicit empirical expression for g, and its differential coefficients we can now obtain explicit expressions for the pressure and density in our model of the universe.

Substituting equations (36) into the expression for pressure given in (34) we obtain

$$8\pi p_0 = -\frac{1}{R^2}e^{-2kt} - 3k^2 + \Lambda. \qquad (38)$$

Owing to the low temperature of intergalactic space, however, it is evident that the pressure that should be ascribed to our simplified model of the universe must be nearly zero over that small time range in the neighborhood of $t = 0$ where we assume our formula valid. This can be obtained from equation (38) only if we have the approximate relation between constants

$$\Lambda = \frac{1}{R^2} + 3k^2 \qquad (39)$$

and we shall introduce this equation below, where it will be seen that its use is equivalent to neglecting the pressure in comparison with the density of the universe at time $t = 0$.

Turning now to the expression for the proper density of matter given by equations (34), and substituting equations (36) we obtain

$$8\pi\rho_0 = \frac{6}{R^2}e^{-2kt} + 12k^2 - 4\Lambda$$

and making use of (39) this can be written in the form

$$8\pi\rho_0 = \frac{6}{R^2}e^{-2kt} - \frac{4}{R^2}. \qquad (40)$$

With the help of this expression connecting density and time, we can now obtain an expression for the rate at which matter is changing over into radiation. For the proper volume dVo associated with the coordinate range $dx\ dy\ dz$ we can evidently write, in accordance with the line element (37)

$$dV_0 = \frac{e^{3kt}}{\left(1+\frac{r^2}{4R^2}\right)^3}dx\,dy\,dz. \qquad (41)$$

Hence using equation (40) for the proper density of matter, we can write

for the mass of matter M included within a given range of coördinates x, y, z

$$M = A\left\{\frac{6}{R^2}e^{kt} - \frac{4}{R^2}e^{3kt}\right\} \tag{42}$$

where the quantity A, whose magnitude depends on the range of coordinates taken, is a constant independent of the time. This mass, however, which is included within a given coördinate range, can only change with the time through the transformation of matter into radiation, since our model of the universe has been so constructed that material particles are stationary in the coördinates used and thus do not cross any coördinate boundary. Hence differentiating equation (42) with respect to the time, and dividing by M itself, we easily obtain for the percentage rate of the transformation of matter into radiation at the time $t = 0$ the very simple expression

$$-\left[\frac{1}{M}\frac{dM}{dt}\right]_{t=0} = 3k. \tag{43}$$

Equation (43) gives us an intermediate interpretation for the quantity k in terms of the average rate at which matter is disappearing throughout the universe.[10]

§ 9. *Relation of the Quantity k to the Red Shift in the Light from Distant Objects.*—It is also possible to obtain an interpretation for the quantity k in terms of the red shift in the light from the nebulae.

In accordance with the general theory of relativity, we can secure an expression for the velocity of light by setting ds equal to zero in our formula for the line element. Doing so and restricting ourselves to distances where r is small compared with R, we obtain for the velocity of light

$$\frac{dl}{dt} = e^{-kt}. \tag{44}$$

The restriction to small distance r introduces considerable simplification, since it limits our considerations to a region in which space is approximately Euclidean so that the coördinate distance l can easily be correlated with the distance as measured by usual methods. Some restriction is in any case necessary, however, since our formula for the line element can certainly claim validity only for times in the neighborhood of $t = 0$, and hence could not be applied with confidence if the light required too long a time to reach the observer.

Consider now a light impulse which leaves a nebula at the time t_1 and reaches the observer at time t_2, after traveling the distance Δl as measured in our system of coördinates. By integrating equation (44) we then obtain

$$\Delta l = \int_{t_1}^{t_2} e^{-kt} dt = \frac{1}{k}\left(e^{-kt_1} - e^{-kt_2}\right). \tag{45}$$

Since the nebula will be stationary in our system of coördinates, we obtain on differentiating this expression with respect to the time of departure t_1

$$0 = \frac{1}{k}\left(-e^{-kt_1} k + e^{-kt_2} k \frac{dt_2}{dt_1}\right)$$

or solving for dt_2/dt_1, we can write

$$\frac{dt_2}{dt_1} = e^{k(t_2 - t_1)}. \tag{46}$$

We have thus obtained an expression connecting the short time interval dt_1 between the departure of two light impulses from the nebula with the time interval dt_2 between their reception. And considering dt_1 as the period of the light on emission and dt_2 its period on reception we can at once write the relation

$$\frac{\lambda + \delta\lambda}{\lambda} = e^{k(t_2 - t_1)}, \tag{47}$$

where $\delta\lambda$ is the shift in wave-length of the light from the nebula as compared with that from a similar terrestrial source.

Moreover, considering times t_1 and t_2 in the neighborhood of $t = 0$, which in any case is necessary in order to accord with our assumptions, this formula for the change in wave-length can be rewritten in a more familiar form, since by expanding (45) and (47) we obtain the approximate relations

$$\Delta l \approx t_2 - t_1 \quad \text{and} \quad \frac{\delta\lambda}{\lambda} \approx k(t_2 - t_1)$$

and by combining can write

$$\frac{\delta\lambda}{\lambda} = k\ \Delta l. \tag{48}$$

We have thus secured another interpretation of the quantity k, as the constant which occurs in the empirically verified linear formula connecting red shift with distance.

§ 10. *Comparison of the Magnitudes of k as Determined from the Red Shift and from the Rate of Transformation of Matter into Radiation.*—We must now compare the magnitude of k, as determined from observations on the red shift in the light from the extra-galactic nebulae, with the values that would correspond to the rates of transformation of matter into radiation occurring in the case of stars of different types.

Expressing the distance Δl to the nebulae in light years, it is found

that equation (48) fits the experimental data of Hubble and Humason[4] on the change of red shift with distance with an approximate value of the constant

$$k = 5.1 \times 10^{-10} \text{ (years)}^{-1}. \qquad (49)$$

On the other hand for the rate of transformation of matter into radiation in the stars we obtain a wide range of values depending on the type of star considered. The rates for a series of typical stars are given in the table below, together with the corresponding values of k as calculated from equation (43). The first two columns of the table are taken from Jeans, *Astronomy and Cosmogony*, page 125.

GENERATION OF ENERGY BY TYPICAL STARS

STAR	ERGS PER GRAM PER SECOND	$-\frac{1}{M}\frac{dM}{dt}$ PER YEAR	k YEARS^{-1}
H. D. 1337 A	15,000	5.2×10^{-10}	1.7×10^{-10}
B. D. 6° 1309 A	(11,000)	3.9×10^{-10}	1.3×10^{-10}
V Puppis A	1,100	3.9×10^{-11}	1.3×10^{-11}
Betelgeux	(300)	1.1×10^{-11}	3.7×10^{-12}
Capella A	48	1.7×10^{-12}	6.3×10^{-13}
Sirius A	29	1.0×10^{-12}	3.3×10^{-13}
Sun	1.90	6.6×10^{-14}	2.2×10^{-14}
α Centauri B	0.90	3.2×10^{-14}	1.1×10^{-14}
60 Kruger B	0.02	7.0×10^{-16}	2.3×10^{-16}

It will thus be seen that the value, $k = 5.1 \times 10^{-10}$ years^{-1}, determined from the red shift falls close to the range of values that would be determined from known rates for the change of matter into radiation. Nevertheless, the value from the red shift certainly seems high when account is taken of the preponderance of stars of the type of the sun or later, and we must return to this again when we try to estimate the validity of the suggested explanation of the red shift.

§ 11. *Conclusion.*—In conclusion, I wish to make several comments concerning different features of the line element that we have obtained. I also desire to make some statement as to what degree of probability we could feel justified in ascribing to this tentative explanation of the red shift.

a. Relation to the Einstein Line Element.—A close relationship will be noted between the line element which we have obtained in this work—

$$ds^2 = -\frac{e^{2kt}}{\left(1 + \frac{r^2}{4R^2}\right)^2}(dx^2 + dy^2 + dz^2) + dt^2 \qquad (50)$$

and that of Einstein, since by an easy transformation of coördinates we can change his line element from the usual form given by equation (1), § 1, to the form[11]

$$ds^2 = -\frac{1}{\left(1+\frac{r^2}{4R^2}\right)^2}(dx^2+dy^2+dz^2)+dt^2, \qquad (51)$$

which bears a close relation to our line element. And on the other hand, by the reverse transformation we could, if desired, re-express our line element in a form more similar to the usual Einstein one.

The quantity k, which occurs in the term that distinguishes our line element from the Einstein one, is proportional to the average rate at which matter is changing into radiation, and if we should take this to be zero, our line element would reduce to that of Einstein. Hence with considerable correctness we can describe the model of the universe which corresponds to our line element as a *non-static* Einstein universe—necessarily non-static since it takes into account a general process of transformation of matter into radiation neglected in the derivation of the older formula.

b. Relation to the de Sitter Line Element.—Some measure of relationship can also be noted between our line element and that of de Sitter, since Robertson[5] has shown by his transformation of coördinates that the de Sitter line element can be written in the form

$$ds^2 = -e^{2kt}(dx^2+dy^2+dz^2)+dt^2 \qquad (52)$$

which appears to have considerable relation to ours.

For several reasons, however, I think that too much stress should not be laid on the similarity between expressions (50) and (52).

In the first place, our line element has been derived on the definite hypothesis of a universe containing matter which is gradually being transformed into radiation, while the de Sitter line element can only be derived by definitely neglecting the effect of any matter which actually may be present in the universe.

In the second place, the term involving the time t enters our equations because of our hypothesis of the transformation of matter into radiation, and the sign and magnitude which we obtain for the constant k from a consideration of known rates of transformation accounts correctly for the sign and approximately for the order of magnitude of the observed red shift in the light from distant objects. On the other hand, in the case of the de Sitter line element in the form given by equation (52), we should have no *a priori* reason which would guide us in choosing either the sign or magnitude of the constant k. Either sign would lead to agreement with the gravitational equations for an empty universe, and in the absence of knowledge as to the magnitude of the red shift, any reasonably small value of k would seem equally possible.

Finally, it should be noted that the term involving the time enters the

equation (52) in the exact way demanded by the de Sitter hypothesis of a substantially empty universe, while it is presumable that the term involving the time in our equation (50) gives only a first approximation of a more correct form that would correspond to a more exact fitting of our equation to the changing rate at which matter is beng transformed into radiation. Indeed, it is almost certain that the form that we have taken is only a first approximation, since by carrying out a double differentiation of equation (42) it can be shown that our form leads, at time $t = 0$, to a rate of transformation of matter into radiation which is increasing with the time, and this result does not appear probable.

c. The Recession of the Nebulae.—The system of coördinates used in this article was chosen in such a way that particles having zero coördinate velocity would continue to have no motion in these coördinates. From the form of our line element (50) it is evident, however, that the distance from the origin to a nebula permanently located at $r = \sqrt{x^2 + y^2 + z^2}$ *when measured with rigid scales* would have the value

$$\int ds_0 = e^{kt} \int_0^r \frac{dr}{1 + \frac{r^2}{4R^2}} \tag{53}$$

and hence be increasing with the time. Thus, in this sense and in others that will be evident, we can speak of a motion of recession of the nebulae, and can think of the red shift as correlated with this motion. Such a recessional motion, however, is not to be regarded as an arbitrary device for accounting for the red shift, but as a necessary accompaniment of the change in gravitational field which is produced by the transformation of matter into radiation.

d. The Validity of the Explanation.—Finally, a few remarks which attempt to estimate the validity of this attempted explanation of the red shift will not be out of place.

First of all, it should be clearly recognized that the explanation is definitely founded on the hypothesis of a continuous and uncompensated process taking place throughout the universe; and the explanation would certainly fall to the ground, if in reality the universe should prove to be in a steady state, the mass of the stars being continuously replenished by some cyclical process whose steps are now unknown. There is, indeed, little evidence in favor of such a cycle and astrophysicists are not inclined to this view; nevertheless I myself and many others would be glad to give it credence if we could.

In the second place, account must be taken of the simplifying assumptions that have been made in the course of the argument. These have been of two kinds—the mathematical approximations that have been introduced in taking only two terms of our empirical formula (35) for the

dependence of g on t, and the high degree of over-simplification introduced into the model of the universe that we have used for our calculations. I do not think that the effect of the mathematical approximations is necessarily negligible since the observational determination of the connection between red shift and distance has perhaps not yet been pushed far enough to discover deviations from a linear relation. On the other hand the over-simplification of the model may have considerable effect on the results. The model that we have used for our calculations is essentially that of a universe filled with a homogeneous distribution of matter and radiation with a continuous increase in the radiation at the expense of the matter. In the actual universe, however, at least a large portion of the matter is really concentrated in stellar systems, and in addition to the transformation of matter into radiation which seems to be taking place inside these systems, there appears over a long time range to be an excess in the outflow of radiation accompanied by a cooling and readjustment in the structure of the systems. At a later time I shall perhaps be able to carry out the computations for a less unduly simplified model and further investigate the effect of the mathematical approximations.

In the third place, we must not forget the fact that the value of k determined from the red shift, although falling close to the range that corresponds to the actual rates of transformation of matter into radiation found in the stars of different types, is nevertheless considerably higher than would be expected in view of the large preponderance of stars of the type of the sun or later. This is certainly a serious difficulty for the theory, although not necessarily a fatal objection until we have made calculations using less approximation and simplification and have obtained more certain information concerning the rates of dissipation of mass in the nuclei of the nebulae, in dust which may be present throughout the whole of intergalactic space, and possibly also in those as yet unknown processes which are the cause of the cosmic rays.

In conclusion, we can at least say that in order to get a satisfactory line element for the universe as a whole we must certainly take into account the possibility of general evolutionary processes taking place throughout the universe. And the non-static line element which we thus obtain may provide the explanation for the whole or a part of the red shift in the light from the extra-galactic nebulae.

[1] Einstein, *Berl. Ber.*, 1917, p. 142.

[2] de Sitter, *Monthly Notices, R. A. S.*, **76, 77, 78,** 1916–1917.

[3] Einstein's original derivation contained an equivalent to the assumption that the pressure of this fluid was zero. This is not necessary, however, see Tolman, *Proc. Nat. Acad. Sci.*, **14,** 350 (1928), and reference 6 below.

[4] See Hubble, *Proc. Nat. Acad. Sci.*, **15,** 168 (1929); and Humason, *Ibid*, **15,** 167 (1929).

[5] Robertson, *Phil. Mag.*, **5**, 835 (1928).

[6] Tolman, *Proc. Nat. Acad. Sci.*, **15**, 297 (1929).

[7] Robertson, *Ibid.*, **15**, 822 (1929).

[8] Zwicky, *Ibid.*, **15**, 773 (1929).

[9] For a solution giving the first integral of this equation I am indebted to my colleague, Professor Paul S. Epstein, whom I wish to thank also in this place.

[10] Two remarks can be made which may make it easier to understand the treatment given in this Section. (1) The symbol ρ_0 signifies the proper density as measured by an observer actually moving with the material at the point in question. It hence applies to matter alone, as the rest mass of a light quantum is zero; and since matter is stationary in our coördinates it gives the desired proper density of matter in those coordinates. (2) Noting that the macroscopic density of radiation is three times its pressure, it can be shown from the equations of this section that the rate of decrease of material energy in a given coördinate range is compensated for by the rate of increase of radiation in that range.

[11] Einstein, *Berl. Ber.*, 1918, p. 448.